

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:18 ; Search time 25.55 Seconds
(without alignments)
5154.826 Million cell updates/sec

Title: PCT-US01-13059-2
Perfect score: 9089
Sequence: 1 MQSIMDSSAVNATEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3949	43.4	1017	T48452	hypothetical prote
2	2863	31.5	555	T48453	hypothetical prote
3	1559.5	17.2	1207	D84781	hypothetical prote
4	1260	13.9	234	T48454	hypothetical prote
5	709	7.8	917	T05430	hypothetical prote
6	278	3.1	5327	T13564	microtubule-associ
7	246.5	2.7	2897	B48666	cell proliferation
8	246.5	2.7	3256	A48666	cell proliferation
9	240.5	2.6	2938	T30249	cell proliferation
10	233.5	2.6	2218	B84683	hypothetical prote
11	232	2.6	259	D75275	endonuclease III -
12	219.5	2.4	1871	D96796	probable heat shoc
13	217.5	2.4	1490	T20513	hypothetical prote
14	217	2.4	2022	T48818	glucan 1,4-alpha-g
15	211	2.3	1188	T05324	hypothetical prote
16	210.5	2.3	1805	A34736	nestin - rat
17	208.5	2.3	3924	S37431	ankyrin 2, neurona
18	207	2.3	2331	T25410	hypothetical prote
19	206	2.3	3187	JC5837	364K Golgi complex
20	205.5	2.3	2562	T14266	xin protein - chic
21	205	2.3	1983	T00385	KIAA0624 protein -
22	204.5	2.2	2717	A34203	DNA-binding protei
23	197.5	2.2	1145	T18235	transcription acti
24	197	2.2	1366	B86292	hypothetical prote
25	196.5	2.2	3225	T152300	giantin - human
26	196	2.2	3488	T34418	hypothetical prote
27	192.5	2.1	2954	T14156	kinesin-related pr
28	192.5	2.1	3259	A56539	giantin - human
29	192	2.1	1888	T14273	zinc finger protei

30	191.5	2.1	2094	2	S33124	tpr protein - huma
31	191	2.1	1742	2	T49451	kinesin-like prote
32	190.5	2.1	2447	2	T16870	hypothetical prote
33	190.5	2.1	4957	2	T03455	ALR protein - huma
34	190.5	2.1	5262	2	T03454	ALR protein - huma
35	190	2.1	1284	1	WMVZAI	A-type inclusion p
36	190	2.1	2464	1	QRMSP1	microtubule-associ
37	189.5	2.1	236	2	S75373	probable DNA-(apur
38	189.5	2.1	1634	2	T26517	hypothetical prote
39	189.5	2.1	1804	2	T34518	nestin - golden ha
40	189.5	2.1	3498	2	T22330	hypothetical prote
41	189	2.1	2453	2	S60254	nuclear receptor c
42	188.5	2.1	1435	2	A37793	erythrocyte-bindin
43	188.5	2.1	1926	2	JC4842	DNA-binding nuclea
44	188	2.1	1593	2	T22028	hypothetical prote
45	187.5	2.1	264	2	C72770	probable DNA-(apur

ALIGNMENTS

RESULT 1

T48452
hypothetical protein T32M21.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48452
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
Submitted to the Protein Sequence Database, March 2000
A:Reference number: 224487
A:Accession: T48452
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1017 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 167/1; 874/1
A>Note: T32M21.160

Query Match 43.4%; Score 3949; DB 2; Length 1017;
Best Local Similarity 99.9%; Pred. No. 2.3e-206;
Matches 755; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MQSIMDSSAVNATEQNDGSRQDVLEFDLNKTPQCKPSKRRKFKMPKVVVEGPKRKP	60
DB	235	MQSIMDSSAVNATEQNDGSRQDVLEFDLNKTPQCKPSKRRKFKMPKVVVEGPKRKP	294
QY	61	RKPAELPKVVVEGPKRKPRAATQEKVSKETGSAKKNLKEATKKPANVGDMSNKP	120
DB	295	RKPAELPKVVVEGPKRKPRAATQEKVSKETGSAKKNLKEATKKPANVGDMSNKP	354
QY	121	EYTLKSCRKALNFDLENPGDARQGDSEIVQNSGANSFSEIRDAIGTNGSLDVSQ	180
DB	355	EYTLKSCRKALNFDLENPGDARQGDSEIVQNSGANSFSEIRDAIGTNGSLDVSQ	414
QY	181	IDKTGGLGAMNOPLEVSMGNQPKLSTGAKLARDQDPDLTRNQCCQFPVATQTFPME	240
DB	415	IDKTGGLGAMNOPLEVSMGNQPKLSTGAKLARDQDPDLTRNQCCQFPVATQTFPME	474
QY	241	NQAWLQMKNLIGLFFGNQCPRMIRNQCPCLAMGNQCPMYLIGTPRPALVSGNQQLGG	300
DB	475	NQAWLQMKNLIGLFFGNQCPRMIRNQCPCLAMGNQCPMYLIGTPRPALVSGNQQLGG	534
QY	301	PGNKRPIFLNHQTCLPAGNQLYGSPDTHQVLMSTGGQGHGLLIKNQOPGSLIRGQPC	360
DB	535	PGNKRPIFLNHQTCLPAGNQLYGSPDTHQVLMSTGGQGHGLLIKNQOPGSLIRGQPC	594
QY	361	VPLIDQOPATPKGFTHLNQMVATSMSSPGLRPHSQSVPTTYLHVSVSRILNLTGTCCQ	420

Db 595 VPLIDQOPATPKGTHLNMQVATSMSSPGLRPHSQSQVPTTYLHVESVSRILNLTGTCQ 654
Qy 421 RSRAPYDSLOODIHQGNKYILSHEISNGCKKALPONSSSLPTPIAKLEEARSGSRQY 480
Db 655 RSRAPYDSLOODIHQGNKYILSHEISNGCKKALPONSSSLPTPIAKLEEARSGSRQY 714
Qy 481 HRAMGQTEKHDLNLAQQAQSQDYVERHNSSTCVYDLAAAKTKTKQVVQENLHGMPEVI 540
Db 715 HRAMGQTEKHDLNLAQQAQSQDYVERHNSSTCVYDLAAAKTKTKQVVQENLHGMPEVI 774
Qy 541 EIEDDPTDGARKGNTASISKGASGNSSPVKKTAEEKICIVPTPAKKGRAGRKSVPP 600
Db 775 EIEDDPTDGARKGNTASISKGASGNSSPVKKTAEEKICIVPTPAKKGRAGRKSVPP 834
Qy 601 PAHASETOLQOPTPKPTPLSPKPKGRKSIQDSGARGSPGELLGODSTAEIIRYMON 660
Db 835 PAHASETOLQOPTPKPTPLSPKPKGRKSIQDSGARGSPGELLGODSTAEIIRYMON 894
Qy 661 LYLGDKEREQONAMVLYKGDGALVPYVESKRRKPRKVDIDDETRIWNLLMGKDEKEG 720
Db 895 LYLGDKEREQONAMVLYKGDGALVPYVESKRRKPRKVDIDDETRIWNLLMGKDEKEG 954
Qy 721 DEKDKKKKEWEEERRVFRGRADSFARMHLVQGD 756
Db 955 DEKDKKKKEWEEERRVFRGRADSFARMHLVQGE 990
RESULT 2
T48453
hypothetical protein T32M21.170 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48453
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24487
A:Accession: T48453
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-555 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 469/3; 496/2; 524/3
A:Note: T32M21.170

Query Match 31.5%; Score 2863; DB 2; Length 555;
Best Local Similarity 100.0%; Pred. No. 7.8e-148;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 785 SSAPMSLAARPPPKLSSREDERNVRVSVVDDPCCILNLNEIPSWOEKQVHPDMEVSG 844
Db 7 SSAPMSLAARPPPKLSSREDERNVRVSVVDDPCCILNLNEIPSWOEKQVHPDMEVSG 66
Qy 845 VDSGSKQLRDCNSGTERFNFLKSTONLEEVLSQDSFDPALFOSCGRGVSCSCSKS 904
Db 67 VDSGSKQLRDCNSGTERFNFLKSTONLEEVLSQDSFDPALFOSCGRGVSCSCSKS 126
Qy 905 DAEPPTTRCETKTVSGTSQSVQTSQPNLSDEICLQGNRPHLYEGSGDVQKQETTNVAQK 964
Db 127 DAEPPTTRCETKTVSGTSQSVQTSQPNLSDEICLQGNRPHLYEGSGDVQKQETTNVAQK 186
Qy 965 KPDLKTNWKNKDSVCFQGNRNDTNWQTTPTSSSYEQCATRQPHVLDIEDFGMQGGLGYSW 1024
Db 187 KPDLKTNWKNKDSVCFQGNRNDTNWQTTPTSSSYEQCATRQPHVLDIEDFGMQGGLGYSW 246
Qy 1025 MSISPRVDRVKNKNVPRFRFQGGSVPREFTGQIIPSTPHLPGMGLSGSSSAVOEHQDD 1084
Db 247 MSISPRVDRVKNKNVPRFRFQGGSVPREFTGQIIPSTPHLPGMGLSGSSSAVOEHQDD 306

Qy 1085 TOHQNOQEMNKASHLOKTFDLDLNSSECLTRQSTKQNTIDGCLPRDRTAEADVVDPLSN 1144
Db 307 TOHQNOQEMNKASHLOKTFDLDLNSSECLTRQSTKQNTIDGCLPRDRTAEADVVDPLSN 366
Qy 1145 NSSLQNTILVESNNSKKEQTAVEYKETNATILREMGKTLADGKKPTSQWDSLRKQVNEG 1204
Db 367 NSSLQNTILVESNNSKKEQTAVEYKETNATILREMGKTLADGKKPTSQWDSLRKQVNEG 426
Qy 1205 ROERKNKMMDSIDYEARRASISSEISAEIKERGMNMLAVRIKDFLERIVKDHGIDLEW 1264
Db 427 ROERKNKMMDSIDYEARRASISSEISAEIKERGMNMLAVRIKDFLERIVKDHGIDLEW 486
Qy 1265 LRESPPPKAKDYLLSIRGLGLKSVCEVRLTLHLNLAFFVDNNGRIAVRMGVPLQPLPE 1324
Db 487 LRESPPPKAKDYLLSIRGLGLKSVCEVRLTLHLNLAFFVDNNGRIAVRMGVPLQPLPE 546
Qy 1325 SLQHLLEL 1333
Db 547 SLQHLLEL 555
RESULT 3
D84781
hypothetical protein At2g36490 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: D84781
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:2003487
A:Accession: D84781
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1207 <STO>
A:Cross-references: GB:AE002093; NID:g4581149; PIDN:AAD24633.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g36490
A:Map position: 2
Query Match 17.2%; Score 1559.5; DB 2; Length 1207;
Best Local Similarity 28.5%; Pred. No. 1.2e-76;
Matches 480; Conservative 199; Mismatches 366; Indels 641; Gaps 61;
Qy 13 TEATEQNDGSRQDVLEFDLNTPOOKPSKRRKPKMPKVVYEGKPKRPRKPAELPKVYVE 72
Db 65 TEEVESLSSVSNVAE-QILKTP-EKPKRKH--PKVRREAKPKREPAPRPRKSVVD 120
Qy 73 GKPKRPRKKAATOEKVR-SKETGSAKKKLKEKATKPPANVGMNKSPEVTLKSCRKAL 131
Db 121 QGESKTPKRYVRKKVEVSKD-----QDPTVESSAAVETSTRPK---RLCRVL 167
Qy 132 NFDLEN-----PGDARQ-GDSEIVONSSGANSFSEIRDAIGTNGSLDSVSDIKTN 185
Db 168 DFEAENGENTNGDIRAEGMESALQEKLDGSGN-QELKDC-----LSAPTPKRRK 219
Qy 186 GLGAMNPLEVSMGNQDPK-----LSTGAKLARDQDPDLTRN-----QQCQFP 229
Db 220 SQGKRK-----GVQPKNGSNLEEDYISMAQAARKRQGPCTCCDMNLSGTYEQDCDY 271
Qy 230 VATQNTQFPMENQQAQWLQMKNLGIFPFGNQOPRMTIRNQOPCLAMGNQOPMLIGTRP 289
Db 272 -----QKMHLYSPML-----QQGGM--RYDAIC----- 293
Qy 290 ALVSGNOQLGSPQGNKRPIFLNHQTCPLPAGNQLYGSPTDMHQLVMSTGGQOHLIKNQ 349
Db 294 -----SKVFSQQHNVYSAF--HATCYSSSTLSANRV-----LTVEER 331
Qy 350 PGSILRQCPVPLIDQOPATP-----KGFHLNQVATSMSSPGLRPHSQSOV 398

Db 332 EGIFQGRQSELSNVLSDKIDTPIKKTGTHARFNLSSMNKLV-----EV 376
QY 399 PTYLHVESVSRILNGTTTCQSRAPAYDSLQODTHQNKYILSHEISNGNGCKKALPQ 458
Db 377 PE---HL-----TSGYCSKPO-----QNNKILVDTRVT-----401
QY 459 NSSLPPTIMAKLEEARGSKRQYHRAMGQTEKHDLNLAQOIASQDVERINSSTCVEYILDA 518
Db 402 -----VSKKKPTKSEKQ-----414
QY 519 AKKTKIQVVOENLHGMPEVIEIEDPTDGARKGKNTASISGASKNGNSPVKKTAEKE 578
Db 415 ---TK-QKNLNLNLCRFPFSTGLSPD-----437
QY 579 KCIVPKTPAKKGRAGKRSKVPVPAHASEIQLMOPPPKTPLSRSKPKGKGRKSIDSGKA 638
Db 438 -----ELW-----KRRNSI-----446
QY 639 RGPSELLCQDSIAETIYRMQNLXGDKEREQONAMVLYK-----GDGALVPYE 688
Db 447 -----ETISEL-----LRLDIIINREHSETALVPYTNNISQIVLFGGAGAIVPVT 490
QY 689 S-KKKRPKPVDDIDTTRINLLMGKDEKKEKKEKKEEERVRFRGRADSEI 747
Db 491 PVKKPRPKVLDLDDTDRWKLLL-ENINSEGVDSDEQAKAKWEEERVRFRGRADSEI 549
QY 748 ARMLVQGRDRRSPKNGSVVDSVIGVFLTONVSDHLSSAFMSLAARFP-PKLSSSREDE 806
Db 550 ARMLVQGRDRRTPKNGSVVDSVIGVFLTONVSDHLSSAFMSLAARFPVFPVPSNFD- 608
QY 807 RNVRSVVVEDPEGCIILNLEISWQ-----EKVQHPDSMEVSGVDSSKQLRDCSNS 859
Db 609 -----AGTSSMPSIQITYLDSSETSSPPDHNHSSVT-----640
QY 860 GIERFNFLKSIQNLSEELVSSQSDPDPAIFQSCGRVSGSCSKSDAEPPTTRCET--KT 917
Db 641 -----LANTQDEKDVPSNET-----SRSSSEIAISAHESVDKT 676
QY 918 VSGTS--QSVQGTGSPNLSDEICLOGNERPHLYEGSGDVOK-QETTINVAOKKPDLEKTMW 974
Db 677 TDSKEVVDSDRGKS-----SVEVDKTEKCRVLNLFPSSEDSALTC 716
QY 975 KDSVCGFQPRNDNTNQTTSSVEQCATRQPHVLDIEDFGMGEGLGYSWMSISPRVDRV 1034
Db 717 QHSMVSDAPQNT---ERAGSSS-----EIDLEGE-YRTSFMKL-----750
QY 1035 KNKNVPRFRFGGSPREFTCQIIPSTPHELPGMGLSGSSSAVOEHODDTQHNQODENN 1094
Db 751 -----LQGVQVSLSDSNQVSPN-----KSPGDCSSEIKGFQ-----SKKEPT 787
QY 1095 KASHLOKTFDLNLSSEE-CLPQRSSTKONITDGCPLPRDRTAEDVVDPLSNSSNQNLIV 1153
Db 788 KSS-----VDSSEPOCCSQ-----DG-----DVL-----807
QY 1154 ESNSSNKEOTAVEYKETAATILREMGKTLADGKKPTSQWDSLRKQVEGNEGQERKNKMM 1213
Db 808 -----SCQKPT---LKEGKKVILKE-----EKAFDCLDRREAAQARAGIREKTRSTM 852
QY 1214 DSIDYEAIRRASISETSEAIKRGMMNLAVRI-----KDFLERIVKDHGGIDLE 1263
Db 853 DIVDWAIRAADVKEVAETIKSRGNHKLAEIRIQYITLNNKIMQGFGLDRLDVNDHGSIDLE 912
QY 1264 WLRESPPDKAKDYLLSIRGLGLKSECEVRLTLHLNLAFLPVDTNVGRIVARMGWVPLQPLP 1323
Db 913 WLRDVPPDKAK-----923
QY 1324 ESLQLHLELYPVLESIQFLPRLCKLDQRTLYELHYOLITFGVKVCTKSRRPNCNACPM 1383
Db 924 -----YELHYQMITFGVKVCTKSRRPNCNACPM 950
QY 1384 RGECHRFASAYASARIALPAPERISLT SATIPVPPSPFPVPAIMLEPLPLEKSLASCA 1443
Db 951 KGECHRFASAFASARIALPSTEGMGTDPKNPLP-----LHLPEPTFOREGGSEV 999

QY 1444 PSNREN-----CEPIIEPASPQGECTEITESDIEDAYVNEDEPDPIPTIKLINIEQFGMT 1497
Db 1000 VQHSEPAKVKTCCETPIIEPASPPEETAETAEVSIADIEAEFF-EDPEIPTIRLNMDAFTSN 1058
QY 1498 LREHMERNELOEGDMSKALVALHPTTTTIPPKLNISRLRTEHQVYELPDHRLDGM 1557
Db 1059 LKKIMEHNKELQDGNSSALVALTAETASLPMPKLNISQLRTEHRVYELPDHPLLAQL 1118
QY 1558 DKREPDPSPYLLAIWTPGETANSAPPPQKCGKASGKMGCFDETCSECSNLSREANSQT 1617
Db 1119 EKREPDPSYLLAIWTPGETADSIQPSYSTCIFQANGMLCDEETCFSCNSIKETRISQIV 1178
QY 1618 RGTLLI 1523
Db 1179 RGTILV 1184

RESULT 4

T48454
hypothetical protein T32M21.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48454
R:Bevan, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24487
A:Accession: T48454
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 49/2; 61/3; 81/1; 127/3; 149/3; 188/1; 201/1
A:Note: T32M21.180

Query Match 13.9%; Score 1260; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 2e-61;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1496 MTLREHMERNMELQEGDMSKALVALHPTTTTIPPKLNISRLRTEHQVYELPDHRLD 1555
Db 1 MTLREHMERNMELQEGDMSKALVALHPTTTTIPPKLNISRLRTEHQVYELPDHRLD 60
QY 1556 GMDKREPDPSPYLLAIWTPGETANSAPPPQKCGKASGKMGCFDETCSECSNLSREANSQ 1615
Db 61 GMDKREPDPSPYLLAIWTPGETANSAPPPQKCGKASGKMGCFDETCSECSNLSREANSQ 120
QY 1616 TVRGTLIPCRTRAMRGSFPLNGTYFOVNEFLADHESLKPIDVPRDWIWDLPRTTYFGT 1675
Db 121 TVRGTLIPCRTRAMRGSFPLNGTYFOVNEFLADHESLKPIDVPRDWIWDLPRTTYFGT 180
QY 1676 SVTSIFRGLSTEIQICFWKGFVGVVGFQKTRAPRPLMARLHFFPASKLKNKT 1729
Db 181 SVTSIFRGLSTEIQICFWKGFVGVVGFQKTRAPRPLMARLHFFPASKLKNKT 234

RESULT 5

T05430
hypothetical protein F28A23.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
C:Accession: T05430
R:Bevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.; Herzl, A.
submitted to the Protein Sequence Database, October 1998
A:Reference number: Z15415
A:Accession: T05430
A:Molecule type: DNA
A:Residues: 1-917 <BEV>

A;Cross-references: EMBL:AL021961

A;Experimental source: cultivar Columbia; BAC clone F28A23

C;Genetics:

A;Map position: 4

A;Introns: 417/3; 70/3; 88/1; 118/1; 321/1; 367/1; 474/1; 504/2; 624/3; 655/2; 667/3; 695/3

A;Note: F28A23.180

C;Superfamily: Arabidopsis thaliana hypothetical protein F28A23.180

Query Match 7.88; Score 709; DB 2; Length 917;
Best Local Similarity 21.48; Pred. No. 1.1e-30;
Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;

```
QY 651 IAEIYRMQNLVLDGEREQEONAMVLYKGDGALVPYESKRRKPRKVDIDDETTRIWNL 710
DB 388 IAKLIKDMGRKINKK-----VTTMIKADKLV-----TAKVLDPEIKEDV 431

QY 711 LMGKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEK 770
DB 432 LM--VNDSPSRSYDDKETEAKWKEREIFQTRIDLFNRHRLQGNKFKQWKSGVVDV 489

QY 771 IGVFELTONVSHLSSAFMSLAARFPKLSRSSREDERNVRSVVVEDPEGCILNLEIPSW 830
DB 490 VGVFELTONVSHLSSAFMSLAARFPKLSRSSREDERNVRSVVVEDPEGCILNLEIPSW 830

QY 831 QEKVQHPDMEVSGVSGSKEQLRDCNSGIERPNFLEKSTONLEEVLSODSFDAIF 890
DB 531 -----DAKSECI----- 538

QY 891 QSCGVSGSCSKSDAEPTTRCTKTVSGTSQSVQVGSNLSDEICLOGNERPHLYEGS 950
DB 539 -----ILSDE----- 543

QY 951 GDVQKQETTVAOKKPDLEKTMNKKDSVCFGCPNDTNWQTTSSSYEQCATROPHVLDI 1010
DB 544 ----- 543

QY 1011 EDFGMOGELGYSMWSISPRVDRYKKNVPRFRFGSGVPRFTGQIIPSTHELPGMG 1070
DB 544 ----- 543

QY 1071 LSGSSAVQEHODTQHNQOQEMNKASHLOKTFDLNLSSECLTROSSYKQNTDGCCLP 1130
DB 544 ---SISKVEDH-----TAKRNEKTGII- 565

QY 1131 RDRFAEDVDPVLSNNLSQNLVNESNNKEQTAVEYKETNATILREMGKTLADGKKPTS 1190
DB 566 -----EDEIVD----- 571

QY 1191 QWDSLRLKDVEGNEGRQBRNNNDSDYEAIRRASISEIAKERGMNNMLAVRIKDFL 1250
DB 572 -WNNLR-MYTKEG--SRPEHMDSVNNSVRLSGQVLETTIKRGQFRILS----- 620

QY 1251 ERIVKHGGIDLEWLRSPDPKADYLLSIRGLGLKSEVCRLLTLNLNAPPVDVTNVGRI 1310
DB 621 ERIL-----VDNTNVGRI 632

QY 1311 AVRMGWVLOPLPESLOHLELYPVLESIQFLWPLRLKLDORTLYELHYQLITFGKVF 1370
DB 633 AVRLGLVPLEPLNGVQNH-----QFELYELHYQLITFGKVF 669

QY 1371 CTRSRPCNACPMRGCRHCFASAYASARLAPAEERSLTSATIPVPPESF-----PPVA 1425
DB 670 CTKTIPNCNACPMKSECKYFASAYVSVKLVLESPEEK-----HEPNFMNASHODVA 722

QY 1426 IPMIELPLPLEKSLASGAPNRCENPEIIEPASPQOECTEITESDIEDAY---NEDPD 1482
DB 723 VDMTSNINLVEECVSGSCSQQAICYPLVFPFSSPRAEIPEST--DIEDVPFNNLYQSYA 780

QY 1483 EIPTKILNIEQFGMTLREHM--ERNMELOGDMSKALVALHPHTTSTPTP---KLKNISR 1537
DB 781 SVPKIDFDLALKSVEDALVISGRSSSDEEISKALVIPTPENACIPKPPKMKMYNR 840
```

```
QY 1538 LRTEHQVYELPDSHRLLDGMDKREPDDPSPYLLAIWTPGETANSAOPPEQKCGKASGKM 1597
DB 841 LRTEHVYVVLDPDNHLLHDFERRKLDPSYLLAIWQ----- 878

QY 1598 CFDETCSECNLSREANSQTVRGTLILPCRTAMRGSPFLNGTYFOVNFELFADHESSLPID 1657
DB 879 ----- 878

QY 1658 VPRDWINDLPRTVYFGTSVTSIFRGLSTEQJCFKMGVFCVGRGFEQKTRAPRLMARL 1717
DB 879 -----GFLCLRAFDRKORDPKELVRL 900

QY 1718 HFP 1720
DB 901 HTP 903

RESULT 6
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N;Alternate names: hypothetical protein EG:49E4.1
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13564
R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17689
A;Accession: T13564
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-5327 <SPA>
A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0025392
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A;Note: EG:49E4.1
C;Superfamily: Drosophila 576K microtubule-associated protein homolog
```

Query Match 3.18; Score 278; DB 2; Length 5327;
Best Local Similarity 19.7%; Pred. No. 3.3e-06;
Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;

```
QY 2 QSTMDSAVNATEQNDGSRDVLFDLNTKTPQOKPKRRKFMKPVVVEG-----K 55
DB 1986 ESKD-----EAKSKEESRRE-----SVAEKSPLSKEASRPASVAESIKDAEK 2031

QY 56 PKRPRKPAELPKVYVEGPKRKP-----RKAATQEKYKSKETGSAKKNNKESATKK 108
DB 2032 SKEESRRESVAEKSPLSKEASRPASVAESIKDAEKSKESRRESAAEKSPLSKEASR 2091

QY 109 PANVG-----DMSNKSPEVTLKSKRKALNFLENPGDAR--QGDSE--SEIVQNSGANSF 160
DB 2092 PASVAESVKDEADKSKESRRE-----SMAESGKAOSIKGDQSPLEKVSRSPEVAESV 2144

QY 161 SEIRDAITGGTNGFSDVSQIDKTNGLGAMNPOKLSGTGAKLARDQOQDILL 220
DB 2145 KD--DPVKSEPRRESVAGSVTADSDQSPLESGKARPSVVDVSDKAEKQESRR 2202

QY 221 TRNQOQCFPVATONTOPMENQOAWLQMKNLQIGFFEGNOQPRMTIRNQOQCLAMGNOQP 280
DB 2203 ESKTESVIPPKAQDKDPSKEVLQ-----PVSMTETIREDADQ- 2240

QY 281 MYLIGTPRALVSGNQO-----LGGPQGNKRPFLNHOITCLPAGNQLYGSTDMHQLV 333
DB 2241 M-----KPSQAESRRESIAESIKASPRDEKSPLEKASKEASRPGSVAESIKYDLDKPKII 2294

QY 334 MSTGGQOH-----GLLTKN-----QOQSLRGQOQPCVPLID--OOPATPK-----GF 374
DB 2295 KDKSTEHRSRELEDKSAVTSKSVSRPLSVASDHEAAVAIEDDAKSSISPKDKSRPGF 2354

QY 375 THLNQMVATSMSSGRLRPHRSQSQVPTTYLHVSVSR-----ILNGTTGTCQRSRAPAYD 428
```



```
Db 2355 -----VAETVSP-----IEATWESKIEVVEKSSIALSLQGGSGKLTODSSPV-- 2400
Qy 429 SIQODIHQNYLSHSHEISNGCKKALPQNSLPTPIMAKLEEARGSKRQYHRAMGOTE 488
Db 2401 ----DVAEGD---PSHAVA-----SVSTVTPTLTKPAELA-----QIGAATV 2436
Qy 489 KIDLNAQOIAQS-QDVERHNSSTCEVILDAAKTKIKQVVOENLHGMPEVIEIDD-- 545
Db 2437 SSPLEALRTSPAPHSIHASDPAESACASIEIASODKSPQVLKES--SRPAWAEKDDAA 2494
Qy 546 ----PTDGARKGNASTSKGASKGN--SSPVKKT-----AEKECIVKPTPAKKGRA 592
Db 2495 QLKSSVEDLRSPVASTETSRPASAGETASSPIEEAPKDFAEFEQAEKAVLPITIELGNL 2554
Qy 593 GRKKSVPVPAHASEIQLWQTPPKTPLSRKPKGKRKSIODSGRKARGPSGELLQODSTA 652
Db 2555 PTLSSPVDVAHAS-VQPAELSKVDIEKTASSPIDEAPKSLIGSPAEERPESPAESAKDAA 2613
Qy 653 EIIYRMQNLILGDKEREQONAWL-----YKDGALVPYESKRRKPRPKVDIDDETT 705
Db 2614 ESV-----EKSADSRPSPSVVESTKADSTKGDISPSP-ESVLEGPDKDDVEKSKSS 2663
Qy 706 RIWNLMLGKDEKDEKDKKKEKWEERVRFRGRADSFITARMHLVQGDRRFSPWKS 765
Db 2664 RPPSV-----SASITGDSTKDSR-----PAS 2685
Qy 766 VYDSVIGVFLTONVSHLSSAFMSLAARFPFKLSSREDERNVRSVVVEDPEGCILNLN 825
Db 2686 VVESV-----KDEH-----DKAESRRESIAKVESVI--DEAG----- 2715
Qy 826 ELPQWQKQVQHPDMVEVSGVSGSKEQRLDSCNSGIERNFLEKSIONLE-----EVL 879
Db 2716 -----KDSKSSQSDQKDEKSTLASKEASRESVSVSSKDDAEKSESRESPI 2764
Qy 880 SS-----ODSPDPAIFQSCGRVGC-----SCSKSDAEFPTRCETKIVSGTS 922
Db 2765 ASGEVPRESKPLDSKOTSRPGSVVESVTADEKSEQOOSRESVAESVKADTKK-DGKS 2823
Qy 923 QSVQTSPLNLSDEICLOGNE-----RPHLYEGSGDVQKQETTNAQKKPDLEKTMNWKDSV 978
Db 2824 Q--EASRPSSVDELLKDDDEKQESRRQSTGSHKAMSTMGDESPPMDKADKSEPSRPESV 2881
Qy 979 CFGQPRNDTNQWTPSSSYEQCATQPHVLD-IDFQMGQEGGLGYSWMSIS-PR--VDRV 1034
Db 2882 AESIKHENTKDBESPLGS-----RDSVAESIKSDITKGEKSPSPKSEVSRPESVWGS 2935
Qy 1035 KKNVPFRFRFGGVSVPREFTGQII---PSTPHELPGHGLSGSSSAVQEHQDDTQHNOOD 1091
Db 2936 KDEKAESRESVAESVKPESSKDATSAPPSKEHSRP-----ESVLGSLKD 2980
Qy 1092 EMNKASHLQKTFDLNLNSEECLTRQSSTK-----QNTIDGCLPRDRTA-----EDVVD 1140
Db 2981 EGDKTTSRVSVADSIKDEKSLVLSQEAASRSEAEKSLDAAAPSQETSRRPESVTVESVKD 3040
Qy 1141 PLSNNSSQLNILLVESNSNKEQTAVEYKETTATILREMK-GTLADGKPTSQWDSLRKDV 1199
Db 3041 GKSPVASKEASRPASVAENAKDSADESKQRPESLPQSKAGSIKDEKSPASKDAEAKSK 3100
Qy 1200 EGNEGROERNKNMDSIDYEAIRRASISEI--SEAIKERGNNML-----AVRIK 1247
Db 3101 E--ESRRESVAEQFPLVSKEVSRPASVAESVKDEAKSEKSEPLMSKEASRPASVAGSVK 3158
Qy 1248 DFLERIVKDHGIDLEWLRSEPPDKA 1273
Db 3159 DEAEKSKE-----ESRRESVAESK 3177
```

RESULT 7

B48666

cell proliferation antigen Ki-67, short form - human

C:Species: Homo sapiens (man)

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999

C:Accession: B48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u ins.
A:Reference number: A48666; MUID:94043435
A:Accession: B48666
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SCH>
A:Cross-references: EMBL:X65551
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.7%; Score 246.5; DB 2; Length 2897;
Best Local Similarity 18.8%; Pred. No. 7e-05;
Matches 362; Conservative 266; Mismatches 768; Indels 527; Gaps 94;

```
Qy 22 SRQDVLEFDLNTPOOKPSKRKRKMPKVVEGK-----PKRKR--KPAELPKVVVEGKP 75
Db 1137 SQDPVDVDTPTSSKPSKRSR-----KVDVEEFFALRKRTSPSAGKAMHTPKPAVSGEK 1190
Qy 76 KRKPKAATQEKVKSKETCSAKKNL---KESA-----TKKPAVGDMSNK 118
Db 1191 NIYAFMGTPVQKLDLTENTLGTSKRRLQTPKEKAQALDLAGFKELFQTRGHTESMTNDK 1250
Qy 119 SPEVLTKCRKALNFLENPGDARGDSESEIVQSSNGANSFSEIRDAIG---GYNQ--- 172
Db 1251 TAKVACKSQPDLD---KNPASSKRR-----LKTSLGKVGKVEALLAVGKLTQTSGETT 1301
Qy 173 -----SFLDSVQI-DKTNGLGAMNOPLEVSMGNQ---PDKLSTGAKLARDQ 215
Db 1302 HTHTPTGDKGSMKAFMESPKOILDSASLTGSKRLQTPKGRTPKCKSEVPEDLAGFIELF--Q 1359
Qy 216 QPDLITRNQOCQFPVATQNTQFPMENQO-----AWLQMNQLIGFFPGNQ-QPRMTIR--- 267
Db 1360 TP-----SHTKESMTNEKTKVSYRASQPDLDVDTPTSSKPKRSLRKAD 1404
Qy 268 NOOPLMAGNQQPM--YLIGTPRALVSGNQLGGPGQGNKRPIFLNHOTCLPAGNQLVGS 325
Db 1405 TEEFLAFRKKQTPSAGKAMHTPKPAV-----GEEKDI-----NTFLGT 1442
Qy 326 PTDMHOLVMTSGGQGHLLIKNQPGSL--IRG-----QQPCV--PLIDQQPATPKGFTH 376
Db 1443 PVOKLDQPNLPGSNRRLOTRKEKAQALELTGFRELFTQCTDNPDADEK-----TT 1495
Qy 377 LNMVATSMSSPGLRPHSOSQVPTTYLHVESV-----RILNGITGTCQSRAPAYDSL 430
Db 1496 KKILCKSPQSDPADTPTNTKQPKRSLKKADYEEBFLAFRKLTPSAGKAMHTPKPAVGP- 1554
Qy 431 QODIHO--GNKYILSHSISNGNGCKKALPQNSLPTPIMAKLEEARGSKRQY-----HRA 484
Db 1555 EKDINTFVGTPEKLDLIGNLPGSKRR-PQT---PKEKAKALEDLGKFKELQTFCHTEE 1610
Qy 485 GQTEKHLNLAQOIAQSODVERHNSSTCEVILDAAKTKIKQVVOENLHGMPEVIEIED 544
Db 1611 SMTDDKITEVSCSKSPQDPVKPTTSS-----KQRLKISLQKV-----GVKEEVLV 1656
Qy 545 DPTDGARKGNASTSKGA-----SKGNSSPVKTKAEKECIV-----PKTPA 587
Db 1657 -----GKLTQTSKGTQTHRETAGDGKSIKAFKESAKQMLDPANYGTGTMERPTPK 1708
Qy 588 KKGR-----AGRKKSVPVPAHASE-----IQLWQTP-----PKTPLSR 621
Db 1709 EEAQSLDLGKFKELFQTPDHTTEESTTDDKTKIACKSPPPESMDTPTSTRRRPTPLGK 1768
Qy 622 -----SKPKG---KGRKSIQDSGK-----ARGPSGE 644
Db 1769 RDIVEELSALKQLTQTTHTDKVPGDEKGINVRETAKQKLDPAASVTSKSKRQPRTPK 1828
Qy 645 LLCQDSIAEIIYRMQNLVLDGKEREQONAMVLYKGD-----GALVPYESKRRKPRPKVD 699
```

```

Db 1829 AQLPDLGLAGLKFQTPVCTDKTTHKTKIACRSQPDVPVGTPTIFKQSKRSRKAD 1888
QY 700 IDDETRINLLMGKDEKE- ---GDEKDKKKKEWEEERRVFRGRADSFIAHML-- 752
Db 1889 VEESIALRRTPSVGKAMDTPKPAGDEKDKM- ---AFMGTP- ---VOKLDLPG 1935
QY 753 -VQDRRFSPW- ---KGSVDHVSIVG- ---VFLTNVNSDHLSSAFMSLAAREP- --- 796
Db 1936 NLPGSKR- ---WPQTPREKAQALDLAGFKLFQTPGTDKTTHKTKIACKSPQDPD 1992
QY 797 PKLSSREDNRNVSVVDEPGCIILNLEIPSWQEKVQHPDSMEVSGVDSGKEQLRDC 856
Db 1993 TPASTKQRPKNRLKADVE- ---EFLALRKRTPSAGKAMDT- ---KPAVSDE 2038
QY 857 SNSGIERFNFLEKSIONLE- ---EVLSSQDSFDPFAIFQSCG- --- 894
Db 2039 KNIN- ---TEVTFVOKLDLGLNLPKSGKROPQTPKEKAQALDVLGF- KELFQTPGHTEE 2093
QY 895 - ---RVGSCSKSDAE- EPTTRCE- ---TKVSGTSQSQVOTG 928
Db 2094 SMTDDKITEVSCSKSQPQESFKTSRSSKQRLKIPLVKVDKKEEPLAVSKLRTSGETQTH 2153
QY 929 SPNLSDBIQCG- NERP- HLYEGSDV- ---OKQETTNNVQAKPKDLEKTMNWKDSVCFQGR 984
Db 2154 TEPTGDSKSIKAFKESPKQIOLDPAASVTGSRRLQTRKKEKARALEDLVDPKE- LFSAPG 2211
QY 985 NDTNWTTPSSSYEOCATROPHVLDIEDFCMGQGLGYSNWSISPRVDRVKNKNVPRRF 1044
Db 2212 HTESMTIDKNTKIPCKSPPELTD- ---ATSKKCPKTR- ---PRKEV 2253
QY 1045 ROGGSVPREFTGQIIPST- PHELPGMLSGSSSAVOB- ---HODDTQHNQDQEMNK 1095
Db 2254 KEELSAVERLTQSGSTHTKPEPASGDEGIVLKQRAKKKPNVEEPPRRRPAPKEK 2313
QY 1096 ASHLQKT- ---FLDLNLS- ---EGLTROSSTKQNTDGLCPDRDRTAEDVDVPLNNSSLQN 1150
Db 2314 AQLPDLGLAGLKFQTPVCTDKTTHKTKIACRSQPDVPVGTPTIFKQSKRSRKAD 1888
QY 1151 ILVESNSNKEOTAVEYKFNAT- ---LREMGKTLADGKKTPOWDSLRKDEGNE 1203
Db 2367 TRVQKVQVEEPAVFTQSGTETDADKEPAGEDGKGA- LKESAKQTPAPAASTGSR 2425
QY 1204 GROERNKNNMDSIDYFAIRASISEIAIKERGMNMLAVRIKDFLERIVKHGGIDLE 1263
Db 2426 RRPAPRESAQATE- ---DLAGEKDPAAAGHTESMTDDKTKIPCKSSPELDTATSS- 2479
QY 1264 WLRESPPDKA- ---KDYLLSIRGLGLKSVCEVRLLLTLHLNLAFTVDTNVGRIAVRWGP 1318
Db 2480 - ---RRPRTRAQKVEVEELAVGLTQSGE- ---THTDKPEVGEKGKTKAFK- 2527
QY 1319 LQPLPESLQHLLELVPVLESIQFLWPRKLDQRTLYEL- ---HYQLITFG 1367
Db 2528 -QAKRN- ---VDAEDVIGRRQ- ---PRAPKEKADLELASFQELSQTPGHTEELANG 2578
QY 1368 KVFCTKSRPNCNACPMRGCRHFPASAYASARLALPAPEER- ---SLTSATIPVPPSPFPV 1424
Db 2579 AADSETSAP- ---KQTPDSGK- ---PLKISRRVLRAPKVEPVGDVYSTRDPVKSQSKNT 2630
QY 1425 AIPMIELPLPLEK- ---SLASGAPSNRECEP- ---IIEB- PASPGQECTEITSDIEDAY 1476
Db 2631 SLP- ---PLPFRGGGKDGSVTQTKRLRCMPAPEEIVEELPASKKOR- ---VAPRA 2678
QY 1477 YNBDPEIPTKINIEQFGMTLREHMERNMELQGDMSKALVALHPTTTSIPTPK- --- 1531
Db 2679 RGSSPVPVIMKSLR- ---TSKRIPEAEELNSNDW- KTNKEEHLQDVSVPENKGLSLR 2733
QY 1532 LKNISRLRTEHQVIEL- PDSHRLDGMKREDDPDPSPYLLAIWTPGETA- --- 1579
Db 2734 SRQDKTEAEQITEVFVLAERIEINRNEKKPMKTSPE- MDIQNPDDGARKPIPRDKVTE 2792
QY 1580 - ---NSAQPP- FQKCGGKASGMKCFD- ---ETCSECSLREANSQTVR 1618

```

```

Db 2793 NKRLRSARQNESSQPKVAEESGGQSAKVLQMONKQKGEAGNSDSMCLRSRRTKSQAP 2852
QY 1619 GTL 1621
Db 2853 STL 2855
RESULT 8
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
J:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge
J:Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u
ins.
A:Reference number: A48666; MUID:94043435
A:Accession: A48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: EMBL:X65550; NID:g415818; PIDN:CAA46519.1; PID:g415819
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.7%; Score 246.5; DB 2; Length 3256;
Best Local Similarity 18.8%; Pred. No. 8.3e-05;
Matches 362; Conservative 266; Mismatches 768; Indels 527; Gaps 94;

QY 22 SRODLVEFLDKTPOOKPSKRRKFKMPKVVEGK- ---PKRKP- ---KPAELPKVWVEGKP 75
Db 1496 SQDPDVPDTPSSKPSQKRSR- ---KVDVEEFFALRKRTPSAGKAMHTPKPAVSGEK 1549
QY 76 KRPRKAATQEKYKSETGSAKKKNL- ---KESA- ---TKKPAVNGDMSNK 118
Db 1550 NIYAFMGTPVOKLDLENLTGSKRRRLQTPKEAQALEDLAGFKELFOTRGHTESMTNDK 1609
QY 119 SPEVTLKSKKALNFLENPGDARQDSESEIVQNSGANSFSEIRDAIG- ---GTNG- --- 172
Db 1610 TAKVACKSQPDLD- ---KNPASSKRR- ---LKTSLGKVGKVEELAVGLTQSGETT 1660
QY 173 - ---SFLDSVSQI- DKTNGLGAMNQPLEVSMGNO- ---PDKLSTGAKLARQ 215
Db 1661 HTTEPTGDKSKMKAPEWSPKQILDNSAALTGSKRQLRTPKGSVEPDLGAFLELF- Q 1718
QY 216 QPDLLFRNQOCQPPVATQNTQFPMENQ- ---AWLQMKNLQIGFPFGNQ- QPRMTIR- --- 267
Db 1719 TP- ---SHTKESMTNEKTKVSYRASQPDVDTPTSSKPPKRSRLKAD 1763
QY 268 NQOPCLAMNQOPM- ---YLICTPRPALVSGNQQLGGQGNKRPFLAHQTCPLPAGNOLYS 325
Db 1764 TEEFLAFRQKQTPSAGKAMHTPKPAV- ---GEEKDI- ---NTFLGT 1801
QY 326 PTDMHOLVMTSGQQHGLIKNOQPGSL- ---IRG- ---QOPCV- ---PLIDQOPATPKGFTH 376
Db 1802 PVQKLOQPGNLPGSNRLQTRKEKAQALELTGTFELFOTPCDTNDTADK- ---TT 1854
QY 377 LNMVATSMSSGLRPHSOSQVPTTYLHVSVS- ---RILNGTTGTCORSRAPYDSL 430
Db 1855 KKLCKSPQSDPADTPTNTKQRPKRSLLKRAADVEEFLAFKRLTPSAGKAMHTPKAAVGE- 1913
QY 431 QODIHO- ---GNKYLLSHEISNGCKKALQNSLPTPIMAKLEEARGSKROY- ---HRAM 484
Db 1914 EKDNFTVGTPTVEKLDLGLNLPKSKRR- POT- ---PKEAKALEDLGAFKELFOTPGHTE 1969
QY 485 GOTEKHDNLNAQOIAQSQDVVERHNSSTCVLEYLDAAKTKTKIQVQENLHGMPPPEVTEIED 544
Db 1970 SMTDDKITEVSCSKSQPDVPKPTSS- ---KQRLKISLGKV- ---GVKEEVLVP- 2015
QY 545 DPTDGARKGKNTASISKGA- ---SKGNSSPVKKTAKEKCIIV- ---PKTPA 587

```

```

Db 2016 -----CKLTQTSKTTQTTHRETAGDGSIKAFKESAKOMLDPANYGTGMWRPRTPK 2067
QY 588 KKGR-----AGKKSVPPAHASE-----IQLWQPTP-----PKTPLSR 621
Db 2068 EEAQSLDLAGKELFQTPDHTTESTDKTKIACKSPPPESMDTPTSTRRPKTPLOK 2127
QY 622 -----SKPKG---KGRKSQDSGK-----ARGPSGE 644
Db 2128 RDIVELSALKOLTQTHTDKVPGEDKGINVRETAQKLODPAASVTSKQRPRTPKK 2187
QY 645 LLCQDSIAELIYRMQNLILGDKEREOEQANVLYKGD-----GALVPYESKRRKPRPKVD 599
Db 2188 AOPLEDLAGLKELFQTPVCTDTPKTHEKTKIACKSPQDPVGTPTIFKPSKRSRLRAD 2247
QY 700 IDEETRIWNLMLGKDEKE-----GDEKDKKKKEWEEERRVFRGRADSFIAHML-- 752
Db 2248 VEEESIALRKRTPSVGKAMDTKPKAGGDEKDKM-----AFWGTGTP---VQKLDLPG 2294
QY 753 -VOGDRRFSW-----KGSVDSVVG---VFLTQNVSDHLSSAFMSLAARFP----- 796
Db 2295 NLPGSKR---WPTPKAKAQALEDLAGKELFQTPGTDKPTTDEKTKIACKSPQDPVD 2351
QY 797 PKLSSREDERNVRSVVEDPCCILNLNEIPSWQEKVQHPDMEVSGVDSKSKQELRDC 856
Db 2352 TPASTKQRPKNLRKADVE---EEFLALRKRTPSAGKAMDTP-----KPAVSDE 2397
QY 857 SNSGIERFNLEKSIQNLSE-----EEVLSSQDSFDPAPFQSCG--- 894
Db 2398 KNIN---TFVETPVQKLDLLGNLPGSKRQPTPKREKAEALDLYGF-KELFQTPGHTEE 2452
QY 895 -----RVGSCSKSDAE-FPTTRCE-----TKTVSTGTSQSVQRG 928
Db 2453 SMTDDKITEVSKSQPESFKTSRSQKRLKIPLVKVDMKEEPLAVSKLRTSGETTQTH 2512
QY 929 SPNLSDICLOG-NERP-HLYEGSDV---QKQETTNVQAKKPDLEKTMNKKDSVCFQGR 984
Db 2513 TPTGDSKSIKAFKSPKQILDPAASVTSGRKQLRTRKEKARALEDLVDKE--LFSAPG 2570
QY 985 NDTNMQTTPSSYEQACRPHVLDTIEDFGMOGELGYSWMSISPRVDRVKNKNPRRPF 1044
Db 2571 HTEESMTIDKNTKIPCKSPPELTD-----ATSTKRCPKTR---PRKEV 2612
QY 1045 RGGGVPRFTQGITPST-PHELPQMLGSGSSAVQJ-----HQDQTQNHQDQEMK 1095
Db 2613 KEELSAVERLTQTSQSTHTEKAPASGDEGIVLKQRAKKKPNPVEEESRRRPRAPKEK 2672
QY 1096 ASHLQKT--FLDLLNSS---EECLTRQSTKQNTDGLCLPRDRTAEDVDVPLSNSSLNQ 1150
Db 2673 AOPLEDLAGLTSEFSGHTQESLTAGKATK-----IPCESPPLEVVDTTASTKRHLR 2725
QY 1151 ILVESNSSNKEQTAVEYKETNATI-----LREMKGTLADGKKPTSQWDSLRKQVEGNE 1203
Db 2726 TRVQKVQVKEEPSAVKFTQSTGETTDADKEPAGEDKGIKA-LKESAKQTPAPAAVTSGR 2784
QY 1204 GRQERNKNMDSIDYBAIRASISEISEAIKRGMMNMLAVRIKQPLERIVKDHGIDLE 1263
Db 2785 RRPAPRASAQA-----DLAGFKDPAAGHTBESMTDDTKTKIPCKSSPELDTATSS-- 2838
QY 1264 WLRESPPDKA---KDYLLSIRGLGLKSEVCVRLTLNLNLAFFVDVTVNVRTAVRMGWVP 1318
Db 2839 --KRRPRTRAKVEKVEELAVCKLTQTSGE-----TTHDKEPVGEGKGTAKF----- 2886
QY 1319 LQPLPESLQHLLEYPVLESIQKFLWPLRCKLDORTLYEL-----HYQLITRG 1367
Db 2887 -QPAKEN---VDAEDVIGSRQ---PRAPKEKAQPLEDLASFQBSLQTPGHTTELANG 2937
QY 1368 KVFTCKSRNCNACMRGCRHFASFASAYASARLALPAPEER---SLTSATIPVPPESFPV 1424
Db 2938 AADSEFTSAP--KQTPDSGR-----PLKTSRRVLRAKPEVPEDVTVSTRDPVKSQSKNT 2989
QY 1425 ATPMIELPLEK---SLASGAPSNRECEP---LIEE-PASPGQECTEITESDIEDAY 1476

```

```

Db 2990 SLP-----PLPFRKGGKDGSVTGTTRKLRCMPAPEEIVEELPASKKQR-----VAPRA 3037
QY 1477 YNEDDEIPTIKLNTIEQFQMTLREHMERNMELQEGDMSKALVALHTTTSIPTPK----- 1531
Db 3038 RGSSEPPVIMKRSR---TSAKRIEPAEELNSNDM-KTNKEEHKLQDSVPENKGISLR 3092
QY 1532 LKNISRLRTEHQVYEL-PDSHRLLDGMKREDPPSPYLLAIWTPQETA----- 1579
Db 3093 SRRQDKTEAQOQITVEFVLAERIEINRNKPKMKTSP-MDIQNPDDGARKPIPRDKVTE 3151
QY 1580 -----NSAOPP-EQKCGKSGKMGCFD-----ETCSECNLSREANSQTVR 1618
Db 3152 NKRCLRSARQNESSQPKVAESGGOKSAKVLQMONQKQKGEAGNSDMLSRKTKSQPAA 3211
QY 1619 GTL 1621
Db 3212 STL 3214

RESULT 9
T30249
cell proliferation antigen Ki-67 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30249
R:Starborg, M.; Gell, K.; Brundell, E.; Hoog, C.
J. Cell Sci. 109, 143-153, 1996
A:Title: The murine Ki-67 cell proliferation antigen accumulates in the nucleolar and
i for cell cycle progression.
A:Reference number: 220787; MUID:96431717
A:Accession: T30249
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2938 <STA>
A:Cross-references: EMBL:X82786; NID:g1177527; PIDN:CAA58026.1; PID:g1177528
A:Experimental source: strain CBA; testis
C:Genetics:
A:Gene: Ki-67
C:Keywords: cell cycle control; nucleus; tandem repeat

Query Match 2.6%; Score 240.5; DB 2; Length 2938;
Best Local Similarity 18.4%; Pred. No. 0.00015;
Matches 363; Conservative 266; Mismatches 699; Indels 643; Gaps 91;

QY 33 KTPQOK-----PSKRRKRFMPVYVVEGKPKRPAELPKVYVVEGKP---KRPKR 81
Db 1175 ETPKQKLESIEIENLGLRKQSRTPKDTIGFQDSFIQIPDHANGPLVYVVKTKKMFNSQPES 1234
QY 82 AATQEKVKSKETGSAAKKNLKESATKKPANVG-----DMSNKSPEVTLSKRCALNFDLE 136
Db 1235 AITRKSRRQSRASISKIDVKEELLESEHLQLGEGVDTFQVSTNKVIRSSRKAAPKRLD 1294
QY 137 N---PGDARQDSESEIVQNSSGANSFSEIRDAIGTNGSFLDSVS----- 179
Db 1295 STAGMPSNMRKSCSKDNTPCLEDLNGFQELFQMPGYANDSLTGTSTMLARSPQLGPVR 1354
QY 180 -QIDKTN-----GIGAM-----NOPLEYSMGNQPKLST--- 207.
Db 1355 TQINKSLPKIILRKMDVTEISGLWKQSLGRVHTTQEQEDNAIKAIMETPKTLQTAAD 1414
QY 208 GAKLARDQOPDLLTRNQOQFPVATONTQFPMENQOAWLQ-----KNOLIGFFPGNQ 260
Db 1415 GTRLTRQPO---TPKEVQ-----PLEDHSVFQELFQTSRYCSDPLI---GNK 1456
QY 261 QPRMTIRNQOQCLAMGNQOQPMVILGTPRALVSGNQOLGPGQGNKR--PIFL-----NHQ 313
Db 1457 QTRMSLRSPQ-----GFVTRPTSKRLAKTSVGNIAVREKISPSLSLQCATGEV 1506
QY 314 TGLPAG-----NQLYGSPTDMHQLVMSTGG-----QQHGLLIKNOQPSGLIRGQPCVPL 363
Db 1507 VHIPGPDDETENKGVKESTPQTLDSASRTYSKROGAHEERPQFSGDLFHPQE----- 1561

```

QY 364 IDQAPATPKGTHLNMVATSMSPG-----LRPHSQSQVPTTYLH--VESVRLNLTG 417
Db 1562 LFQTPASGKDPVTVDKTKIALQSPQCHIINPASKRQSNMRLKDMREFSILEKQTQS 1621
QY 418 TCQSRAPAYDSLOODLHOGNKYILSHE-----ISNGNCKK--ALPQNSSLPTPTMAK 469
Db 1622 RGRDAGTPA--PMQEE--NGTTAIMEPTKOKLDFIGNSTGHKRRPTPKRAQP----- 1671
QY 470 LEEARGSKROYHRAMGOTEKHDLNLAQOI--AOSO--DVERHNSTCEVEYLDAAKTKIKQ 526
Db 1672 LEDLDGQELFQTPAGASDPVSVESAKISLASSQAEPVPTPAST-----KRSKTKGLSK 1726
QY 527 VVQENLHGMPEVIEIEDDPTDGARKGNKTASISKGASKNSPVKTKAEKCIIVKTP 586
Db 1727 V-----DVRQEPSTLGRMK-----SLGRAPGTPAPVQV--ENDSTAFMETP 1766
QY 587 AKK-----GRAGRKSVPPAHASEIOLWQPTPKTSLRSKPKGKRSIQD-----SGK 637
Db 1767 KQKLDFTGNSGKHRR-----PQPKIRAQP-----LEDLDGQEL 1802
QY 638 ARGPSGELLQDSIAEIIYRMQNLGLDKEREQEQONAMVLYKGDGALVPVYESKK--RKPR 695
Db 1803 FQTPAG--ANDSVT-----VEESVKMSLESSQAEPVKTPTASTKRLSKTGL 1845
QY 696 PKVIDDITRIWNLMLMGKDEKDEKDEKKEKKEWEEERRVFRGRADSFAR----- 749
Db 1846 SKVDVREDPSILEKTKTSPTGTPAPVQVQENDCTAFMETPKOKLDFGTGNSGKHRRPTPKI 1905
QY 750 ---MHLVQGDRRSPKGSVVDVIGVFLTONVSHLSSAFNSLAARPPKLSRSSRED 805
Db 1906 RAQPLELDGQELFQTPAGASDSV-----TVEESAKMSLESSQAEPVKTPTASTKRLS 1958
QY 806 ERNVRV--VVEDPEG--ILNLNEIPSWQEKVOHPSD-----MEVSGVDSGSK 850
Db 1959 KTGSLKVDVREDPTSLGKTKSPGRACGTPAPVQVQENDCTAFMETPKOKLDFGTGNSGSK 2018
QY 851 EQLRDCNSGIERENFLEKSIQNLLEEVLSSQSDFAIFQSCGRVSGSCSKSDAEFFT 910
Db 2019 RRSRTSKN---RSQPLE-DLDGQELFQTPAGASNPVSVESAKI---SLESSQAEPVR 2070
QY 911 TRCETKTVSGT---SQSVQVQSPNLSDEICL-----QGN----- 942
Db 2071 TRASTKRLSKTGLNKMVDVREHSPKSLSSCASQKVMQTLTLDGHDGRETGDGKVLAAQL 2130
QY 943 RPHLYEGSGDVQKQETTNVAKPKDL-----EKTWNKDSVCFQGPQNDTNMTPTSS 995
Db 2131 EPAIYVTRGRQQRSCCKRQSPEDLSGVQEVFTSGHNKDSVTV-----DNLAKLPSS 2184
QY 996 SYEQCATRQPHVLIDIEDFGMGEGGLGYSNMSISPRDVRVKNKNVPRFRQGGSVPREFT 1055
Db 2185 S-----PPLETDTSVTSRRQARTGL-----RKVHVKN-----LSGGIMHPQIS 2224
QY 1056 GQIITPSTPHELPGMG--LSGSSSAVQEHODDTQHNQODENMKASHLOKTFDL----- 1106
Db 2225 GEIV-DLPREPEGEGVIKTRKOSVKRKLKTEVNVPRSKRORITRAEKTLEDLPFGQELC 2283
QY 1107 -----LNSSECLTRQSTQONITDGLCPDRDAEDV-- 1138
Db 2284 QAPSLVMSVIVEKTPKMPDKSPPEVDTTSETQARRLRRLRVLTVEEPIDPORKTRVRVQT 2343
QY 1139 ----VDPLSNSSQLNTL-----VESNSNKEQTA- 1164
Db 2344 RNTQKEPISDNQMEEFKESVQKQDPSVSLTGRNRQNPRTVKEKTPQLEELTSFQETAK 2403
QY 1165 -----VEVKETNATILREMGTLA-DGKK--PTQOWDSLKQDVNREGROERNKNNM 1213
Db 2404 RISSKSPQBEKETLAGKRLQIQLINDGKVEPTAQ-----RKQ-PSRETRNTLKEPVG 2458
QY 1214 DSDIYEAIRASISEIAEKRGMMNMLAVRIKDFELRVKDHG-GIDLEWLR----- 1266
Db 2459 DSNINVEEVKSTQKI-----DPVASVPVSKRRPRVPKEKAQALELAGLGPQTLL 2509
QY 1267 -----ESPPDK-----AKDYLLSIRGLGLKSVE 1289

Db 2510 GHTDESASDRGPTQPCNSLQEQVDSFQSRPRTRRGKVEADEPSAVR-----KTVS 2565
QY 1290 CVRLTLTLNLAFVDTNVG-----RIARWGM--VPLQQLPE 1324
Db 2566 TSR-OTMRSRKVFEIGNGTQVSKASTIKQILDVAVKVTGSRRLRTHKGWSTILLKLLGD 2624
QY 1325 SLQI-----HLLLEYPVLESIQK-----FLWP-RLCKLDQRTLYELHYOLITFGKVFC 1371
Db 2625 SKEITQISDSEKLAHDTSLKSTQOQKPDVSKPLRTRCRRVLRSKQVKEVLEVDTRDHA 2684
QY 1372 T-KSRPNCNACPMRGCEHRFASAYASARLAL-----PAPEERSLSATSATIPVPES 1420
Db 2685 TLOSKNPLSPKRSKARDGSIVRTRALRSAPKQEADEKVPPEKRAASSKRYVSP- 2743
QY 1421 FPPVAIPMIELPLPLEKSLAS-----CAPSNRENCEPIIEEPASPGOECEITEITSDIEDA 1475
Db 2744 --PVKMKHLKI---VSNKLESVEBQVSTVMKTEAKRENPTVDDO-----NS 2787
QY 1476 YNEDDPDEIPTIKLINIQFGMTLREHMERNMELQEGDMSKALVALHPTTTSIPTPKL--- 1532
Db 2788 RYRK-----KTNVKQ-----PRKPEDAS 2805
QY 1533 -----KNISRLRTEHQVYELPDHRLDGMKREDDPSYLLAIWTGCTANSQPP 1585
Db 2806 AENVGIKKNEKTKMTASQETEL-----QNPDD-----GAKKTSR-- 2840
QY 1586 EQCGGKASCKMCFDETCSECSILREANSQTVRGCTLLI---PCRTAMRGSP 1634
Db 2841 -----GOVSGK-----RTP-----RGTEMPOPCAEKTSKP 2871
RESULT 10
B84683
hypothetical protein At2g28300 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84683
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: B84683
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2218 <STO>
A:Cross-references: GB:AE002093; NID:g4803953; PIDN:AAD29825.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g28300
A:Map position: 2

Query Match 2.6%; Score 233.5; DB 2; Length 2218;
Best Local Similarity 19.1%; Pred. No. 0.00024;
Matches 345; Conservative 252; Mismatches 644; Indels 563; Gaps 88;

QY 8 SAYNATEATEQDNGSRQDVLEF---DLNKTPOQKPSKRKRKFMKPVVVEGKPK--KRKPRK 62
Db 136 SAYSVRTQATNATSSAATGLDFVSSDKRLEAAHSHTSSSLALTSPD---LSGPPGQFQSLPAS 193
QY 63 PAELPKVVVEGPKRKRKAATQAEKVKSKETGS-----AKKNLKE 103
Db 194 PATP-IRGRGSRGRCGAGRRVGVHGSNSSITQRTETATSLASDAEATKALPR 252
QY 104 SATK-----KPANVGMSNKSPEVTLKSCRKALNFD--LENPGDARQG--DSES-----EI 150
Db 253 SASEIVSRVPKANEGSTSNPDQVSPVHSATTALRSKADKDLADPPGDSGSHVQTLNV 312
QY 151 VQNSSGANSFS-ETDAIGA---TNGSFLLDSVQIDKTNGLGAMNPLFVSMGNQPKL 205
Db 313 LENSERKFAVAKRRLPIQGGGPVQNONVSSVCDGSKSPSEGRTYTALQ-GVTTAP--- 368

Db 198 -----DPPYLHELHNSHGRQVCTWTRPKGKILRERC-----DAYAL 238
QY 1397 ARLALPAPEER 1407
Db 239 YGDKVPSPSEK 249

RESULT 12
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96796
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Nansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: D96796
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1871 <STO>
A:Cross-references: GB:AE005173; NID:g6143906; PIDN:AAF04452.1; GSPDB:GN00141
C:Genetics:
A:Gene: F28016.15
A:Map position: 1

Query Match 2.4%; Score 219.5; DB 2; Length 1871;
Best Local Similarity 17.2%; Pred. No. 0.0011;
Matches 251; Conservative 222; Mismatches 510; Indels 477; Gaps 61;

QY 86 EYKSKETGSAKKNLKESATKPKANVGDMSKSPVTLKSCRKALNF-DLENPDCAROG 144
Db 103 ERTEEEKGLADSNKESVDSLRLP-----PDTEGRECHEQTRHEEQENKQLVQA 152
QY 145 DSESEIVONSSGANSFEIRDAIGTNGSFLDSVSIQIDKTNGLGMNQPLEVSMGN---- 200
Db 153 ESDD---SDDFGSRAFEIEE-----QESDVLDRSTSGAMEKETDMDVGDGLRK 199
QY 201 -----QPKLSTGAKLARDQOPDLITRNQOCQFPVATQNTQFPMENQOAWLMKNQLIGF 255
Db 200 VQIEEPERHNEESKISE-----MVDGETSGHEKKKVVYKM 234
QY 256 PFGNQOQPMRTIRNOOPCLAMGNQOPMYLIGTPRAL-----VSGNQQLGGPQGNKRPIFLN 311
Db 235 DKNRDVKEVDG-----AMGEF-----RPNIDRTQVVGDEDAETAEKNDDEP--- 278
QY 312 HQTCPLAGNOLYGSPTDMHOLVMTSGQGLHLIKNQPSGLINGQOQCPVLIDQOQATP 371
Db 279 -----ESDLEADVOK-----INEGNTK--VRRHSEDRNLKLOEK-----EQHSKEQ 322
QY 372 KGFT---HLNQVATSMSSPCLRHPSQSOVPTTYLVHVESVSRILNGTTTCQSRAPAYD 428
Db 323 KGSKEENMKELV-----EETPT-----EAETIRNDILGPGQIEVPEVD 363
QY 429 SLQODIHQ--NKVILSHSISNGCKKALPQNSSLPTPTMA-----468
Db 364 TLGKTSDEGKEQNIVKELKNGDATEIDAKMGEVFNASNIADTGMNSEDFFESDKLESAD 423
QY 469 ---KLEARGSKQYHRAMQTEKHDNLNLAQITA--QSQDVERHNSSTCVLEYL-----D 517
Db 424 EVDKMWKEDQRENDKVGQASEDISLTKIQLGEIQEQFQCKRHKDKOENIKELREGQASE 483
QY 518 AAK-----KTKLOKVQVENLHGMP-----PEVIE-----IEDDPT----- 547

Db 484 AEKNIKNDILKVPQKRSEGHKHKIOTFQEETNKOPGYNEKIMETGKKINEDGTRKRVQEM 543
QY 548 -----DGARKGKNTASISKGASKGNSSPVKTAETAEKCIKVPKTPAKKGRAGRKSVPP 600
Db 544 IROELDEPARSEKENRSRELKSKTNDKDK-----EKE---IAGTERKERESDRPKILRE 597
QY 601 PAHASEIQLWQTPPKTPLSRSPKPKGKRSIQDSGARGSPGELLQDSIAEIIYRMQ- 659
Db 598 QEVADEV-----AEDTKFSIYGEVEEELAGKEFEFGSDDDIARIVRTEQ 645
QY 660 ---NLYIGDKEREQEQNAMVLYK-----GDGALVPYSEK-----KRPRPKVDIDDE 703
Db 646 LDSNAMQGEQEKMIQELVLEEKVCDGGKGIIVAAETKAENKSKRVQETEEQKLDKEDT 705
QY 704 TTRWNLLMGKDEKGEDEKDKKKKWEERBV-FRGRADSFIAARMHLVQGRDRSPW 762
Db 706 CGHFQKLIIGEISDHGEVEDVEKGRKTAERKIKRDRAREAEIKDKDLGVSGRYI--- 762
QY 763 KGSVVDSDVI---GVFLTQNVSDHLLSSSAFMSLAARFPKLLSSSRREDERNRVSVVVEDEP 819
Db 763 KGTIKELVENRGIYRNEH-----EKKKDDAN-----RPEK 794
QY 820 CILNLEIPSWQEKVQHPSDMEVSGVDSGSK-----EQLRDC-SNSGIE-----RFNLE 868
Db 795 ITGTIK-----QELVSLNSQLROENVEDGDKTQELVEEKIKDCSEEGSSESKIKTDVV 849
QY 869 KSIONLEEVN-----SSQDSFDPALFQSCGRVSCSCSK 903
Db 850 RKVOGIAKEELYPKREHGKTKITELVEETGDYKQEKETAESDIEAECG-----SLRK 904
QY 904 SD-----AEFTTR-----CETKTVSGTSSQVQTGSPN 931
Db 905 VDGIEEHHEPETHKERDNNRVTGAKPSQGEKEKEKIVESMTITENDNSIDVQETK 964
QY 932 LSBEICLOGNERPH-----LYEGSGDVQKQ-----ETTVAOKK----- 965
Db 965 KERPGRLSEHKRYKIQELLMEAGHNDKKEEQNENYTAEVELETERVSSKKVQEGKMD 1024
QY 966 -----PDLEKTMW-----KDSVCFQOPRND-----TNWQ 990
Db 1025 DNSGKFHEFEERKSYEDWTHEKREKRVLYVEEETYPKDKHTGGEDHNDHKEEQKENV 1084
QY 991 TTPSSSYEQCATROPHVLDIEDFGMGEGELGYSWMSTSPRVDRV-----KKNVPRRFRQ 1046
Db 1085 AKAEALNTEEDSFKAKEIEKHODGELKRSVMVQAKROETEEKDTRAMEKNETVERRKQTK 1144
QY 1047 GGSVPREFTGQIIPSTPHELPGMGLSGSSNAVQV-----HODDTQHNOQDMMKASH 1098
Db 1145 DGSU-----GKREGEDPELGGHRRGEEDRIEELVETEISDHKEKVKKDEDIYLSQD 1199
QY 1099 LQKTFDLNLSSEECITRQ---SSTKQNTDGLCPDRDPTAEDVDVPLSNSSLQNLVESN 1156
Db 1200 TKG--VDLGERERRSKORKIHSVEDEIGD--QEDEDAEAAVSVRN-----ENG 1246
QY 1157 SSNKQTAVE-----YKTN-----ATILREMGTLADGKKKPTSQWDSLR 1196
Db 1247 SSRKQVQTEESEKHKQKQKIPETSNPEVNEDESERVVEKETKEVAHVQLEGKTEENC- 1305
QY 1197 KDVEGNEGROERNKNNMDSIDYEAIRASISEIAKERGMNNMLAVRIKDFLIRVKD 1256
Db 1306 KDDGEGRREREKQGMTA-----ENMLRQREK-----TKS 1336
QY 1257 HGGI--DLEWLRESPPDKAK 1274
Db 1337 DDGIVRKIQETKEEPEDEK 1356

RESULT 13

T20513
hypoetical protein F02E9.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T20513

A; Cross-references: EMBL:AL353821; GSPDB:GN00112; NCSP:68B2.20
A; Experimental source: cosmid contig 68B2; strain 74
C; Genetics:
A; Gene: NCSP:68B2.20
A; Map position: 2
A; Introns: 1192/3

Query Match 2.4%; Score 217; DB 2; Length 2022;
Best Local Similarity 17.2%; Pred. No. 0.0017;
Matches 340; Conservative 253; Mismatches 720; Indels 668; Gaps 85;

Qy	35	PQOKSPKRKKFKPVVVVEGKPKRKRKPAELPKVYVVEGKPKRKRKPA	-----ATQEKVK	89
Db	4	PKPKPKD	-----PPSPRPKTKPIITNSKNGAPTISP7KSATALGATPKPSV	50
Qy	90	SKETGSAKKNLKESATKK--PANVGD--MSNKS	-----EVLKSKRALNFDLENPGD	140
Db	51	PTPTNSGLSHHKKPELKKPEPSLLGDFLLGRSPORVAARASAKRRTKMSDAQNVRE	110	
Qy	141	AROGDSESEIV--QNSSGANSEISEIDAIGTNGSFLDSVSQIDKTNGI	-----GAMN	191
Db	111	ELROEMRAAAVRKLIQPGG	-----VRDRVKAMQKASQAAV	160
Qy	192	QPLEVSM--GNOPD	-----KLTGAKLARDQDPDLTRNQOCQFVATQNTQFPMENOO	243
Db	161	EPTFAYNLSDGDDVDEEDRMRIKWRQPKPKPKTIET	-----VEPKGT	208
Qy	244	AWLQMKNOLIGFPFGNOQPRMTIRNOQPCLAMGNQOPMYLIGTTPRALYSGNQLGGPQG	303	
Db	209	STAKEKNKVLIG--RMADDRPDVILKEARP	-----RP--GLPKRIILISDDHWMKKQO	255
Qy	304	NKRPIFLNHOTCLPAGNQLYGSPTDMHQLVMSTGGQOHLIKNQOPGSLIRGOQPCVPL	363	
Db	256	-----KKGKLAKSP	-----VPEAGQ	281
Qy	364	IDQOATPKGFTHLNQMVATSMSPGLRPHSQSOVPTTYLHVESVSRLINGTTGTQCRSR	423	
Db	282	--QNPVSQKKIRDWAORVELPPPPP	-----PVARRPTKTYRHAKT	330
Qy	424	APAYDSLQODIHOGNKYILSHEISNGCKKALPQNSSLPTPIMAKLEARGSKRQYHRA	483	
Db	331	-----DDDHHRG	-----ALSEPDM	351
Qy	484	MGOTEKHDLNLAQIAQSD	-----VERHNSSTCVEYLDAAKTKTIQVQENLHGMPP	537
Db	352	SGESRRSSNPRTQCPNDGIRVRSLOKRGSD	-----DGRISPVKPA	401
Qy	538	EVIEIEDP--TDGARKGNATAS	-----KGAS	573
Db	402	DGIVRPGPPVPSADSSRSRSTVSPSSSGRTPSDRSGSRTPPRGASPPPRRASTPRRA	461	
Qy	574	TAEREKCIVPKTPAKKGRACRKKSV	-----PPPAHASEIOLWQPTTP	615
Db	462	STPLRKASTPKPRASDHSAASDDVIEVPESEVSSKRSPSPPKR	-----RLRSGPPP	517
Qy	616	KTPLRSRK	-----PKGKRKSIOD	634
Db	518	KRLPLRRRGSGGARPKPRRSPPPTTATQTTETTTDDRRPGADKPMPTPRNNGSGSGED	577	
Qy	635	SGKARGPSGELLQODSIAETIY--RMQNLYLGDKEREQEONAMVLYKGDGALVPYESKK	691	
Db	578	SDR--RPPTAAGI--DDLAEITPFGSAFSELEPLRQHTORTA	-----RQS	620
Qy	692	RKPRPKVDIDDETRIWNLL	-----MGKDGKEGDEEKDKKKKEKWEERRV	738
Db	621	AKPKPKQRNESLKLVPNVILKVMYTGAMEKKQEMAEPPRPPPTGKPARIES	-----	673
Qy	739	FRGRADSFIARMHLVQCDREFSPKGSVDVSVIGVLTQNVSDHLSSAFMSLAARFPK	798	
Db	674	LNTVDPFVSGMNLPPVVAPELPLRVSTPE	-----RNSKEKLVDRDLPAHREAPER	725
Qy	799	LSSSREDERNVRSVVYDDEPGGCIILNINETPSWQEKVOHP	-----SDMEVSGV	845

[illegible]

1616 T 1616

Db 1715 S 1715

RESULT 15
T05324
hypothetical protein F1C12.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05324
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15408
A:Accession: T05324
A:Molecule type: DNA
A:Residues: 1-1188 <BEV>
A:Cross-references: EMBL:AL022224
A:Experimental source: cultivar Columbia; BAC clone F1C12
C:Genetics:
A:Map position: 4
A:Introns: 305/3; 348/2; 975/2; 1091/3; 1141/1
A:Note: F1C12.80

Query Match 2.3%; Score 211; DB 2; Length 1188;
Best Local Similarity 17.4%; Pred. No. 0.0016;
Matches 244; Conservative 219; Mismatches 532; Indels 406; Gaps 55;

QY 78 KPRKAATQEKVKSGTSAKKKLNKESATKKPANVGMNKSPEVTLKSKALNF--- 133
DB 20 RPHERTQ---RDNTNNKKK-----NNNNVNIHVAAPERKLNLFVRD 63

QY 134 DLENFG---DARQSGSEIVNGSANGSFSEIRDAIGTNGSFSDVSQIDKTNGLGAM 190
DB 64 HLENCVSVDVDDSIKAVECCSNKSSVSDHR--LSKTESRQDVPS---SSSLGSD 118

QY 191 NOPLEVMGNQDKLSTGAKLARDQPDLLTRNQOQFFVATQNTQFPMENOQAWLOMKN 250
DB 119 SDP---NSGOPENFSRNAASLSVQIWEART-----TOQPPSSNQSLIDSR 161

QY 251 QLIGF-PFGNQPRMTIRNQ---QPLAMGNQPMYLGTPRALVSGNQGLGGPOGNKR 306
DB 162 SMGNSALNSESLESVKESQMIQTEECNNEEBEEIECAPPLESGEKDREGVVRMDI 221

QY 307 PIFLNHQTCLPAGNQLYSGPTDMHOLVMTGGQHGGLLIKNOQPGSLIRGOQCPVPLDQ 366
DB 222 IRKLSNDSETITNDNGSSGNDNSKEVQTEARSPPQVACSPR---IRGQALADLLVQ 277

QY 367 QPA-TPKGFTHLNOAVATSMSSPGLRPHSQSQVPTTYLHVESVSRILNGTTGTCQSRAP 425
DB 278 MTRDRKDLACLREHRCVSKFT-----NRGRIQST-LRIRCYERCLAIQGRHRSKSTA 330

QY 426 AYDSLQDDIHQGNKYILSHEISNGCKKALPONSLPTPIAKLEEARGSKROYHRAMG 485
DB 331 GSDSNRRSGSGVHLLRKY-----KANSNETGTASTSHASTGRIMDKDHPKA-- 381

QY 486 QTEKHDNLNLAQQAQSDVQRHNSSTCVBYLDAAKTK--IQKVVOENLHGMP-----EVI 540
DB 382 -TEKKVL-----QETIEKSGVKEIKVSAVENAKKGVLSISENENGLKLAGETI 432

QY 541 ETEDDPTDQARKGNATASISKASKNSFPVKTAKEKCIVPKTPAKKGRAGRKSVPP 600
DB 433 QKETVEGKRETKREKIVSKESLEKGE--KRESTKKAIAKESVAEKALVG----- 483

QY 601 PAHASEIQWLQPTPPKTPLSRKPGRKSIQDSGKARG----- 640
DB 484 --IAEKVNLW-----NSDEKMRKVVVEKGTGENTINTERVNDVLEATRRI 530

QY 641 -----PSGELLQDSTAEIIRYMN-----LYLGDKEREQEQNAMYLYKGDGA 683
DB 531 LSVESAERSSTTTTSTMTTRCEVAEKVVKKKKDPFVLSLESNKSKEVEDGNI---KPGQV 587

QY 684 LVPYESKKKKRPKVDIDDETTRINLLMGKDEKDEKOKKKKKEWEEERRVFRGRA 743

Db 588 TQOADSCLRRKPAIEEKVLOETT-----VKSDLKKPTKESREERREIEESTNGIA 640
QY 744 DSFIARMHLVQGDREFPWKSGVVDVSVIGVLTQNVYSDHLSSAFMSLAARFPKLSR 803
DB 641 EKVNLWSREKKRRKAMEKKG--GKTEGKAITE-----TNELLQEASR---RISNVE 688

QY 804 EDERNVRSVVVEDPECCIILNLEIPSWQKQVHPDSMEVSGVDSGSKQELRDCSNSGIER 863
DB 689 TAERSIDT-----SWITVVKVGVDRVIMDKRKKSGGETTSAEIGGGK 731

QY 864 FNFLEKSIONLEEVLSOSSDFPAIFQSGRVRGSCSKSDAEFTTTRCETKTVSGTSQ 923
DB 732 EEDL-ASVEAKSKDVIEDRNMPQAVIH-----GS---KERDKE-----RN 768

QY 924 SVQTGSPNLSDEICLQAGNPHLYEGSGDVQKQETT-----NVAQKPDLEKTMNKD 976
DB 769 SSONGEETLS---LRNSEA---KSTKEIERQEVQTEKESVSHGSKREKDKERNSLQYGE 820

QY 977 SVCFGQPRNDTNWQTPSSSYEOCATRQPHVLDIEFGQGEGLGYSWMSISIPVDRVN 1036
DB 821 KMCF--LRNS-----EAKSTKE-----IERNKS 841

QY 1037 KNPVREFFRQGGVPRFTGQIIPSTPHELPGMGLSGSSSAVOEHODDTQHNOODEMNKA 1096
DB 842 QEV-----SQGEESASGHSRESAKENSSQDD-----869

QY 1097 SHLQKTFLLNLSSECLTRQSTKONITDGLPRDRTAEDVVDPLSNNSLQNIIVESN 1156
DB 870 -----ESTHRN-----PNOK--KGIKEPEDE-----ESK 892

QY 1157 SNKEQTAVEYKETNATILREMKGT-----LADGKKPTSQWDSLR 1196
DB 893 KYEREETGENVEEASVEFYNDWDGMEEBEEBEEYGDYFNGDDDHIDISRPRSYWEDLR 952

QY 1197 KDVEGNEGROERNKNMDSIDYEAIRRASISEISEAKERGMNMLAVRIKDFLERI--- 1253
DB 953 KE-----RYLEVLNTESEKK---DICNLIERTVSNFLTSLRQKIDNLMIT 996

QY 1254 -VKDHGGIDLEWLRESPPDKAKDYLLSIRGLGLKSEVCVRLLTLHLNLAFFVDT---NVGR 1309
DB 997 RVQSHGLGVPLNQIEEG-DEYEEWEVECSARNEEDNETEEPEKTNLEAPSDVCSOSSAR 1055

QY 1310 IAVRMGW-----VPLQPLPESLQ---LHLELYPVLESIQKFLWPR 1347
DB 1056 SSTMSWNEFRDQDIDKXNEPTTSLSLPELVPPTNQSTQDMQTISDLKEQMEQLQREM--- 1112

QY 1348 LCKLDQRTL-----YELHYQ 1362
DB 1113 ---LELRNTVKSCIDMQLHFQ 1130

Search completed: July 5, 2001, 12:45:38
Job time: 200 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:18 ; Search time 25.55 Seconds
(without alignments)
5154.826 Million cell updates/sec

Title: PCT-US01-13059-2

Perfect score: 9089
Sequence: 1 MQSIMDSSAVNATEQND.....PRPLMARLHPASKLKNKT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3949	43.4	1017	2 T48452	hypothetical prote
2	2863	31.5	555	2 T48453	hypothetical prote
3	1559.5	17.2	1207	2 D84781	hypothetical prote
4	1260	13.9	234	2 T48454	hypothetical prote
5	709	7.8	917	2 T05430	hypothetical prote
6	278	3.1	5327	2 T13564	microtubule-associ
7	246.5	2.7	2897	2 B48666	cell proliferation
8	246.5	2.7	3256	2 A48666	cell proliferation
9	240.5	2.6	2938	2 T30249	cell proliferation
10	233.5	2.6	2218	2 B4683	hypothetical prote
11	232	2.6	259	2 D75275	endonuclease III -
12	219.5	2.4	1871	2 D96796	probable heat choc
13	217.5	2.4	1490	2 T20513	hypothetical prote
14	217	2.4	2022	2 T48818	glucan 1,4-alpha-g
15	211	2.3	1188	2 T05324	hypothetical prote
16	210.5	2.3	1805	2 A34736	nestin - rat
17	208.5	2.3	3924	2 S37431	ankyrin 2, neurona
18	207	2.3	2331	2 T25410	hypothetical prote
19	206	2.3	3187	2 JC5837	364K Golgi complex
20	205.5	2.3	2562	2 T14266	xin protein - chic
21	205	2.3	1983	2 T00385	KIAA0624 protein -
22	204.5	2.2	2717	2 A34203	DNA-binding protei
23	197.5	2.2	1145	2 T18235	transcription acti
24	197	2.2	1366	2 B86292	hypothetical prote
25	196.5	2.2	3225	2 I53300	giantin - human
26	196	2.2	3488	2 T34418	hypothetical prote
27	192.5	2.1	2954	2 T14156	kinesin-related pr
28	192.5	2.1	3259	1 A56539	giantin - human
29	192	2.1	1888	2 T14273	zinc finger protei

30 191.5 2.1 2094 2 S33124
31 191 2.1 1742 2 T49451
32 190.5 2.1 2447 2 T16870
33 190.5 2.1 4957 2 T03455
34 190.5 2.1 5262 2 T03454
35 190 2.1 1284 1 WMVZAI
36 190 2.1 2464 1 QRMSP1
37 189.5 2.1 236 2 S75373
38 189.5 2.1 1634 2 T26517
39 189.5 2.1 1804 2 T34518
40 189.5 2.1 3498 2 T22330
41 189 2.1 2453 2 S60254
42 188.5 2.1 1435 2 A37793
43 188.5 2.1 1926 2 JC4842
44 188 2.1 1593 2 T22028
45 187.5 2.1 264 2 C72770

ALIGNMENTS

RESULT 1
T48452
hypothetical protein T32M21.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48452
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24487
A:Accession: T48452
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1017 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 167/1; 874/1
A>Note: T32M21.160

Query Match 43.4%; Score 3949; DB 2; Length 1017;
Best Local Similarity 99.9%; Pred. No. 2.3e-206;
Matches 755; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQSIMDSSAVNATEQNDGSRQDVLEFDLNTPOQKPSKRRKMPKVVVEGKPKRP 60
Db 235 MQSIMDSSAVNATEQNDGSRQDVLEFDLNTPOQKPSKRRKMPKVVVEGKPKRP 294
QY 61 RKPAELPKVVVEGKPKRPKPKAATQEKVKSTGSAKKNLKESATKKPANVGDMNSKSP 120
Db 295 RKPAELPKVVVEGKPKRPKPKAATQEKVKSTGSAKKNLKESATKKPANVGDMNSKSP 354
QY 121 EYTLKSCRKALNFDLENPGDARQGDSEIIVNSGANSFSEIRDAIGTNGSLFDSVQ 180
Db 355 EYTLKSCRKALNFDLENPGDARQGDSEIIVNSGANSFSEIRDAIGTNGSLFDSVQ 414
QY 181 IDKTNGLGAMNOPLEVSMGNQDPDKLSTGAKLARDQDPDLITRNQCCQFPVATQTFPME 240
Db 415 IDKTNGLGAMNOPLEVSMGNQDPDKLSTGAKLARDQDPDLITRNQCCQFPVATQTFPME 474
QY 241 NQOAWLQMKNLQIFGFGNQPPRMTRNOOPCLAMGNQOPMYLIGTPREALYSGNQLGG 300
Db 475 NQOAWLQMKNLQIFGFGNQPPRMTRNOOPCLAMGNQOPMYLIGTPREALYSGNQLGG 534
QY 301 POGNKRPIFLNHTQICLPAGNQLYGSPTDMHQLVMTSGGQHQGLLKQKQPSGLIRGQPC 360
Db 535 POGNKRPIFLNHTQICLPAGNQLYGSPTDMHQLVMTSGGQHQGLLKQKQPSGLIRGQPC 594
QY 361 VPLIDQAPATPKGFTHLNMVATSMSSPGLRPHSQSVPTTYLHVESVRIINGTTGTCQ 420
|||||

THIS PAGE BLANK (USPTO)

Db 595 VPLDQDPATKGFTHLNMVATSMSSPGLRPHSOSQVPTTYLHVESVSRILNGTGTQC 654
Qy 421 RSRAPYDLOQDIIHQGNKYILSHEISNGCKKALPONSSILPTIPMAKLEARGSKROY 480
Db 655 RSRAPYDLOQDIIHQGNKYILSHEISNGCKKALPONSSILPTIPMAKLEARGSKROY 714
Qy 481 HRAMGOTEKHLNLAQQAQTAQODVERHNSCTCEYDLDAAKTKTKQVVOENLHGMPPEVI 540
Db 715 HRAMGOTEKHLNLAQQAQTAQODVERHNSCTCEYDLDAAKTKTKQVVOENLHGMPPEVI 774
Qy 541 EIEDDPTDARKGKNTASISGASKGNSPVKKTAEKELVPTPAKKGAGRKKSVP 600
Db 775 EIEDDPTDARKGKNTASISGASKGNSPVKKTAEKELVPTPAKKGAGRKKSVP 834
Qy 601 PAHASEIQLWPTPKTPLSRKPKGKRSIQDSGKARGPSGELLCOQDSTAEIIRYMN 660
Db 835 PAHASEIQLWPTPKTPLSRKPKGKRSIQDSGKARGPSGELLCOQDSTAEIIRYMN 894
Qy 661 LYLGDKEREQSONAMVLYKGDALVPEYSEKKRPRKYVIDIDETTRIWNLLMGKDEKEG 720
Db 895 LYLGDKEREQSONAMVLYKGDALVPEYSEKKRPRKYVIDIDETTRIWNLLMGKDEKEG 954
Qy 721 DEKDQKKKEWEERRVFRGRADSFIAHMLVQCD 756
Db 955 DEKDQKKKEWEERRVFRGRADSFIAHMLVQCD 990

RESULT 2
T48453
hypothetical protein T32M21.170 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48453
R:Bevan, M.; Terry, N.; Argiles, W.; Buysschaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224487
A:Accession: T48453
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-555 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 469/3; 496/2; 524/3
A:Note: T32M21.170

Query Match 31.5%; Score 2863; DB 2; Length 555;
Best Local Similarity 100.0%; Pred. No. 7.8e-148;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 795 SSAFNSLAARPPKLSRSDRENRVSVVDEPPECILNLEIPSWQKQVHPSDMEVSG 844
Db 7 SSAFNSLAARPPKLSRSDRENRVSVVDEPPECILNLEIPSWQKQVHPSDMEVSG 66
Qy 845 VDSGKEOLRDCNSGIERFNFLEKSIONLEEVLSQDSFDPALFQSGRGVSCSCSKS 904
Db 67 VDSGKEOLRDCNSGIERFNFLEKSIONLEEVLSQDSFDPALFQSGRGVSCSCSKS 126
Qy 905 DAEFPTTRCTKTVSGTSQSVQTSQVSPNLSDEICLQGNRPHLYEGSDGVQKQETTNAQK 964
Db 127 DAEFPTTRCTKTVSGTSQSVQTSQVSPNLSDEICLQGNRPHLYEGSDGVQKQETTNAQK 186
Qy 965 KPDLEKTNWQKSVCFQGPDRNTNQTTPSSSYEQCATRQPHVLIDIEFGMQGELGYSW 1024
Db 187 KPDLEKTNWQKSVCFQGPDRNTNQTTPSSSYEQCATRQPHVLIDIEFGMQGELGYSW 246
Qy 1025 MSISPRDVRNKNVPRFRFGGSGVPREFTGQIIPSTPHLPGMGLSGSSSAVOEHODD 1084
Db 247 MSISPRDVRNKNVPRFRFGGSGVPREFTGQIIPSTPHLPGMGLSGSSSAVOEHODD 306

Qy 1085 TOHQDQEMKASHLOKTFDLNLSSEECITROSSTKONITDGCPLRDRTAEDVVDPLSN 1144
Db 307 TOHQDQEMKASHLOKTFDLNLSSEECITROSSTKONITDGCPLRDRTAEDVVDPLSN 366
Qy 1145 NSSLONILVESNSSNKETAVETNATILREMKGTADGKPTSDOWSLRKDVEGNEG 1204
Db 367 NSSLONILVESNSSNKETAVETNATILREMKGTADGKPTSDOWSLRKDVEGNEG 426
Qy 1205 ROENKNNMDSIDYEAIRRASISEISAIRKERNMNNMLAVRIKDFLERIVKDHGIDLEW 1264
Db 427 ROENKNNMDSIDYEAIRRASISEISAIRKERNMNNMLAVRIKDFLERIVKDHGIDLEW 486
Qy 1265 LRESPPDKADYLLSIRGLGLKSVCEVRLTLHLNAPVDTNVGRIAVRMGWVPLQPLPE 1324
Db 487 LRESPPDKADYLLSIRGLGLKSVCEVRLTLHLNAPVDTNVGRIAVRMGWVPLQPLPE 546
Qy 1325 SLQLHLEL 1333
Db 547 SLQLHLEL 555

Query Match 17.2%; Score 1559.5; DB 2; Length 1207;
Best Local Similarity 28.5%; Pred. No. 1.2e-76;
Matches 480; Conservative 199; Mismatches 366; Indels 641; Gaps 61;
Qy 13 TEATEQNDGSRQDVLFDLNTPOQPSKRRKPKVWVVEGPKRPRKPAELPKVWVE 72
Db 65 TEEVESLSSVSNVAAE-QILKTP-EKPKRKH--PKVREARPKRPRKPRKSVVTD 120
Qy 73 GKPKRPRKKAATQEKVK-SKETGSAKKNLKESATKKPANVGMNKSPEVTLKSKRKAL 131
Db 121 QGSKTPARKVVRKVEVSKD-----QDAPTVSSAAVETSTPK---RLCRVL 167
Qy 132 NFDLEN-----PGDARQ-GDSEIVONGSGANSFSEIRDAIGTNGSFLDSVSDIKTN 185
Db 168 DFEAENGENTNGDIRAGEMESALQEKLDGSGN-QELKDCI-----LSAPSTPKRK 219
Qy 186 GLGAMNOPLEVSMGNQDPK-----LSTGAKLARDQDPDLLTRN-----QOCOPF 229
Db 220 SQGKRK-----GVQPKKNGSNLEEVDISMAQAARKRGQPTCCDMNLSGTYDEQCDY- 271
Qy 230 VATQNTQPMENQQAQVLMKNQLIGFPFGNQPPRTIRNQOQCLAMGNQOQPMYLGTPRP 289
Db 272 -----QKHWLYSPNL-----QOQGM--RYDAIC----- 293
Qy 290 ALVSGNQOLGPGQGNKRPILFNHOTCLPAGNQLYSGPTDMHQLVMSTGGQOQHGLLNQK 349
Db 294 -----SKVESQOQHNYSAF--HATCYSSSTQSLSANRV-----LTVEERR 331
Qy 350 PGSILRQOPCVPLIDQOPATP-----KGFTLNQMVATSMSSPGLRPHSOSQV 398

THIS PAGE BLANK (USPTO)

Db	332	EGIFQGRQSELNVLSDKIDTDPDKKKTGTHARFNLMSMKLV	-----EV	376
QY	399	PTTYLHVESVSRILANGTTCQORSAPAYDSLIQODIHQGNKYILSHSISNGCGKALPQ	458	
Db	377	PE-----HL-----TSGYCSKPO-----QNNKILVDTRVT	401	
QY	459	NSSLPTPIWAKLEBARGSKRQYHRAMGOTEKHDLNLAQOIAQODVDEHNSSTCVEYLDA	518	
Db	402	-----YSKKPKTKSEKSO	414	
QY	519	AKKTIQKVVOENLHGMPEVIEIEDDPTDGARKGNKTASISKASKGNSSPVKKTAKE	578	
Db	415	-----TK-QKNLLPMLCRFPSPFTGLSPD	437	
QY	579	KCIVPKTPARKGRAGRKKSVPVPAHASEIQWLQWPTPPKTPLSRSPKKGKRSIQDSGKA	638	
Db	438	-----ELW-----KRNRSI-----	446	
QY	639	RGPSGELLCODSTAEIYYRMQNLYLCDKERQEQBNAMVLYK-----GUGALVPYE	688	
Db	447	-----ETISEL-----LRLLDNRHSETALVPYTMNSQIVLFGGAGAIVPVT	490	
QY	689	S-KKRPRPKVQIDDDTTIRWLLMGKGDEKDEKDKKEKWEERKVRFRGRADFSFI	747	
Db	491	PVAKPRPRKVDLDDTDRTWKLL-ENINSEGVGDSQDEKAKWEERNVFRGRADFSFI	549	
QY	748	ARMHLVQGRFRSPKMGSVYDSVIGVFLTONYSDHLSSAFMSLAARFP-PKLSSSREDE	806	
Db	550	ARMHLVQGRFRPTPKMGSVYDSVVGVELTQNVSDHLSSAFMSLASQFPVPEVPSSNFD-	608	
QY	807	RNVRSVVVEDPEGCIILNLEIPIWQO-----EKVQHPDSMEVSGVDSGSKQLRDCSNS	859	
Db	609	-----AGTSMSPSIQITLYLDSBETMSSPDDHNSSVT-----	640	
QY	860	GIERFNFLEKSIONLEEVLSSODSFDAIFQSCGRVGCSCSKSDAEFFPTRCET--KT	917	
Db	641	-----LANTODEEKDVPSNET-----SRSSSEIAISAHESVDKT	676	
QY	918	VSQGS--QSVQTSNPDLSEICLOGNERPHLYEGSDGVQK-QETNTVAQKKPDLKETMNV	974	
Db	677	TDKSEYVSDRKS-----SVEYDKTDEKRVNLFPSEDSALTC	716	
QY	975	KDSVCFQGRPNDNTWQTTSPSSYEQCATRQPHVLIEDFGMOGBGLGYSWMSISPRVDRV	1034	
Db	717	QHSWSDVAPQNT---ERAGSSS-----EIDLEGE-YRTSEMKL-----	750	
QY	1035	KKNKVPFRFRQGGSVPRFETQIIPSTPHELPMGLMSGSSSAVQEHQDDTOHQOQDEMN	1094	
Db	751	-----LOGVQVSLSDSNQVSPN-----MSPGDCSEIKGFQ-----SMKEPT	787	
QY	1095	KASHLOKTFDLNLSSEE--CLTRQSSSTQNITDGCLPRDRTAEDVDPPLSNSSLQNILV	1153	
Db	788	KSS-----VDSSEPOCCSQQ-----DG-----DVL-----	807	
QY	1154	ESNSSNKEOTAVYEKFTNATILREMKGTTIADCKPKTSQWDSLRLKDXDEGNEGRQERNKMN	1213	
Db	808	-----SCKOPT---LKEKGKVLKE-----EKKAFTDCLRLREAQARAGIREKTRSTM	852	
QY	1214	DSIDYBAIRRASTISETSEAIKRGGMNNLAVRI-----KDFLIRIVKDHGGIDLE	1263	
Db	853	DTVDWKAIRAADVKEVAETIKSRGMNHKLAERIQVLTLLNMKIMOGFDRLVNDHGSIDLE	912	
QY	1264	WLRESPPDKAKDYLILLSIRGLGLSKVECVPLLTLLHNLAFPVDTNVGRTAVRGWVPLQPLP	1323	
Db	913	WLHRDVPDDKAK	933	
QY	1324	ESLQHLLELYPVLESIQKFLWPLRCLKDORTLYELHYQLITFGVKCTKSRPNCNACPM	1383	
Db	924	-----YELHQYMITFGVKCTKSKPNCNACPM	950	
QY	1384	RGECRHFASAYASARLALPAPERSTISATIPVPPSPFPVPAIMIELPLPLEKSLASGA	1443	
Db	951	KGECRHFASAFARLALPSTKMGMTDPKPNLP-----LHLPPEFOEGSEV	999	

[illegible]

THIS PAGE BLANK (USPTO)

A;Cross-references: EMBL:AL021961

A;Experimental source: cultivar Columbia; BAC clone F28A23

C;Genetics:

A;Map position: 4

A;Introns: 41/3; 70/3; 88/1; 118/1; 321/1; 367/1; 474/1; 504/3; 655/2; 667/3; 695

A;Note: F28A23.180

C;Superfamily: Arabidopsis thaliana hypothetical protein F28A23.180

Query Match 7.8; Score 709; DB 2; Length 917;
Best Local Similarity 21.4; Pred. No. 1.1e-30;
Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;

Qy 651 IAEIIRMONLYLGDKEREQONAMVLYGQALVPYESKKRKPRIIDDETRINWL 710
Db 388 IAKLIKDMGLKINK-----VTTMKADKLV-----TAKYNLOPETIKENDV 431
Qy 711 LMKGDEKEDEKDKKEKWEERRVFRGRADSFIAHMLVQDRFSPWKGVSDSV 770
Db 432 LM--VNDSPRSYDDKETEAKWKKEIFQTRIDLFINRHLRLOGNRKFKQWKGVSVDV 489
Qy 771 LGVLTQNVSHLSSAFMSLAAREPPKLSSRDERNVRSVVVEDPSCILNLNEIPSW 830
Db 490 VGVFLTQNTDYLSSNAFMSVAKEPP-----VDAREGLSYIIEEPQ----- 530
Qy 831 QEKVQHPDMEVSGVDSGSKQLRDCNSGIERFNFLEKSIQNLEEVLSQDSFDPAlF 890
Db 531 -----DAKSSCI----- 538
Qy 891 QSCGRVGCSCSKDAEPPTTRCTETKTVSGTSQSVQTSQPNLSDBICLOGNERPHLYEGS 950
Db 539 -----ILSDR----- 543
Qy 951 GDVQKQETTNVAQKPDLEKTMNWKDSVCFGOPRNDTNWQTPSSSYEQCATRQPHVLDI 1010
Db 544 ----- 543
Qy 1011 EDFGMQGEGLGYSWMSISPRVDRVKNKNVPRFRQGGVSPREFTGQIIPSTPHELPGMG 1070
Db 544 ----- 543
Qy 1071 LSGSSAVQEHODTQHNOQDEMKNKASHLQKTFLLNSSECLTRQSTQNTQDGLP 1130
Db 544 ---SISKVEDHEN-----TAKRKNKGTGII- 565
Qy 1131 RDRTAEDVVDPLSNSSLNILVESNSSKKEQTAVYEYKRNATILREMGTLADGCKPTS 1190
Db 566 ---EDEIVD----- 571
Qy 1191 QWDSLRKDVEGNEGRQERNKNMDSIDYEAIRRASISEISEAIKERGMNMLAVRIKDFL 1250
Db 572 -WNLR--MYTKEG--SRPEMHDSVNNSDVRLSGQNVLETTIKRGQFRILS----- 620
Qy 1251 ERIVKHGGIDLEWLRSPDKAKDYLLSIRGLGLKSVCEVRLTLHLNLAFPVDNNGRI 1310
Db 621 ERIL-----VDNNGRI 632
Qy 1311 AVRMGVPLPLPESLQHLLELYLVLESIQKFLWPLRCKLDORTLYELHYLITFGKVF 1370
Db 633 AVRGLVPLPLELPNGVQMH-----QLFEYELHYQMITFGKVF 669
Qy 1371 CTKSPNACPMRCEHRFASAYASARLALPAPERSLTSATIPVPPESF-----PPVA 1425
Db 670 CTKTPNACPMKCEKYFASAYVSSKVLLESPEKM-----HEPNTFMAHSQDVA 722
Qy 1426 IPMIELPLFLKSLAGSPNRENCEPIIEEPASPQCEITEESDIEDAY---NEDPD 1482
Db 723 VDMTSNINLEECVSSGSDQAICYKPLVEFPSSPRAEIPEST--DIEDVPMNLYQSYA 780
Qy 1483 EIPTTKLINEFGMTLRHM--ERNNELOEGDMSKALVALHPTTTSIPTP---KLNLSR 1537
Db 781 SVPKTDFDLDAKSVEDALVISGRMSSDSEISKALVIPTPENACIPIKPPRKMKNYR 840

Qy 1538 LRTEHQVYELPDSHRLDGMKRPDDPSPYLLAIWTPGETANSAPPEQKCGKASGKM 1597
Db 841 LRTEHVYVVLDPNHELLDFERRKLDSPYLLAIWQ----- 878
Qy 1598 CFDETCSECSLRANSQTVRGILLIPCRITAMRGSFPLNGTYFQVNFELFADHESSLAPID 1657
Db 879 ----- 878
Qy 1658 VPRDWINDLPRTVYFCTSVTSIFRGLSTEQIOCFWKGFCVCRGFPQKTRAPRLMARL 1717
Db 879 -----GFLCLRAFDRKQDRPKELVRL 900
Qy 1718 HFP 1720
Db 901 HTP 903
RESULT 6
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N;Alternate names: hypothetical protein EG:49E4.1
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13564
R;Spanos, L.; Papagiannakis, G.; Sidenkiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17689
A;Accession: T13564
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-5327, <SPA>
A;Cross-references: EMBL:AL0311128; PIDN:CAA20006.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0025392
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A;Note: EG:49E4.1
C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.1; Score 278; DB 2; Length 5327;
Best Local Similarity 19.7; Pred. No. 3.3e-06;
Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;

Qy 2 QSIMDSAVNATEATEQNDGSRQDVLEFDLNLKTPQKPSKRKRKMPKVVEG-----K 55
Db 1986 ESIKD-----EAKSKEESRRE-----SVAESPLPSKEASRPASVAESIKDEAEK 2031
Qy 56 PKRPRPAELPKVVVVGKPKRKP-----RKAATQEKVKSKETGSAKKNLKESATKK 108
Db 2032 SKEESRRESVAESPLPSKEASRPASVAESIKDEAKSKEESRRESAAKSPSPSKEASR 2091
Qy 109 PANYG---DMSNKSPEVTLKSCRKALNFLENPGDAR---QGDSE--SEIVONSNGANSF 160
Db 2092 PASVAESVKDEADKSKESRRE-----SWAESGKAQSIKGDQSPKESRPSRPSVAESV 2144
Qy 161 SEIRDAIGGTGSLDSVSQIDKTNGLGAMNQPLEVSGMGNQPKLSTGAKLARQQPDLL 220
Db 2145 KD--DPVKSKEPSRRESVAGSVTADSARDQSPLESKGASRPESVDSVDEAEKQESRR 2202
Qy 221 TRNQOCFFVATQNTQPMENQQAQWLQKMLQIGFPPGNOQPRWTIRNQPCPLANGNOQP 280
Db 2203 ESKTESVTPPKAKDKSPKEVLQ-----PVSMT-----TIREDA-----DQP 2240
Qy 281 MYLGTFRPALVSNQ-----LGGPQGNKRPIFLNHQTCPLPAGNQLYGSPTDMHQLV 333
Db 2241 M-----KPSQAESRRESIAESIKASPSRDEKSPSLASKEASRPGSVASIKYDLDKPQII 2294
Qy 334 MSTGGQOH---GLLKN-----QOPGSLRQGPVCLID--QOPATPK-----GF 374
Db 2295 KDDKSTHERRESLEDKSAVTSKSRPLSVASDHEAANAIVEDADKSSISPDKSRPGF 2354
Qy 375 THLNQVATSMSSPGLRPHSQSQVPTTYLHVESVSR-----ILNGTGTGQSRAPAYD 428

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:17 ; Search time 31.78 Seconds
(without alignments)
3298.265 Million cell updates/sec

Title: PCT-US01-13059-2
Perfect score: 9089
Sequence: 1 MQSIMSSAVNATEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246.5	2.7	3256	21 AAY50976	Human cell cycle p
2	237	2.6	2819	22 AAB35408	Human 07G27 gene
3	210.5	2.3	1805	13 AAR27204	Rat nestin. Rattu
4	210.5	2.3	1805	15 AAR60126	Rat nestin protein
5	208.5	2.3	2432	21 AAY85565	Human homologue of
6	195	2.1	1780	19 AAW53863	Human gravin polyp
7	195	2.1	1780	21 AAB15380	Human gravin prote
8	192.5	2.1	2954	20 AAY01632	Amino acid sequenc
9	191.5	2.1	1786	14 AAR41043	CD4-EBA175 fusion
10	190	2.1	1284	9 AAR81187	Sequence encoded b
11	189	2.1	2453	21 AAB12454	HNRCR protein sequ

12	188.5	2.1	1435	16 AAR70232	P. falciparum SABP
13	188.5	2.1	1435	18 AAW22477	Sialic acid bindin
14	188.5	2.1	1435	21 AAY77900	P. falciparum SABP
15	188.5	2.1	1596	18 AAW31347	Rat tumour suppres
16	188.5	2.1	1604	16 AAR70105	TNF-R-EBA 175 fusi
17	187.5	2.1	1863	21 AAY56028	Human BRCA1 tumour
18	185.5	2.0	1312	21 AAY78884	Retinoblastoma bin
19	185.5	2.0	2409	12 AAR12609	Versican. Homo sa
20	185	2.0	2441	19 AAW40058	Cellular transcrip
21	185	2.0	2441	21 AAY94252	Mouse nuclear CREB
22	184.5	2.0	1346	18 AAW31346	Rat tumour suppres
23	184	2.0	3266	21 AAB42491	Human OREX ORF2255
24	183	2.0	2441	16 AAR79054	CREB binding prote
25	182	2.0	1863	18 AAW26522	Human BRCA1 consen
26	182	2.0	1863	19 AAW79665	BRCA1 (om11) prote
27	182	2.0	1863	19 AAW76100	Human BRCA1 omi3 p
28	182	2.0	1863	19 AAW76098	Human BRCA1 omi1 p
29	182	2.0	1863	21 AAB24217	Human BRCA1 (om11)
30	182	2.0	1863	21 AAB24218	Human BRCA1 (om12)
31	181	2.0	1931	21 AAB23258	Human apoptosis in
32	181	2.0	2781	21 AAY57453	Human transcrip
33	180	2.0	2907	21 AAY57452	Human transcrip
34	179.5	2.0	1863	18 AAW23286	Human breast and o
35	179	2.0	1619	17 AAR81528	BRCA1 mutant from
36	179	2.0	1663	15 AAR46608	Plasmodium falcipa
37	178.5	2.0	778	21 AAG46504	Arabidopsis thalia
38	177.5	2.0	3248	17 AAR99795	Kinetochore protei
39	177	1.9	1863	21 AAY77818	BRCA1 protein sequ
40	176.5	1.9	1484	20 AAW89721	Canine ribosome re
41	176.5	1.9	1978	20 AAY07032	Breast cancer asso
42	175	1.9	1210	16 AAR66450	AF-4 protein (enco
43	174.5	1.9	2101	21 AAY49936	Human Numa protein
44	174	1.9	1863	17 AAR81536	BRCA1 mutant from
45	173	1.9	1863	16 AAR76641	BRCA1 protein. Ho

ALIGNMENTS

RESULT 1
AAY50976
ID AAY50976 standard; Protein; 3256 AA.
XX
AC AAY50976;
XX
DT 10-MAR-2000 (first entry)
XX
DE Human cell cycle protein Ki-67.
XX
KW Cell cycle protein; Ki-67; therapy; cell proliferation; allergy;
KW tumor treatment; autoimmune disease; scar formation; inflammation;
KW rheumatic disease; transplantation.
XX
OS Homo sapiens.
XX
PN DE19822954-A1.
XX
PD 25-NOV-1999.
XX
PF 22-MAY-1998; 98DE-1022954.
XX
PR 22-MAY-1998; 98DE-1022954.
XX
PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
XX
PI Flad H, Gerdas J, Boehle A, Deinert I;
XX
DR WPI; 2000-039964/04.
XX
DR N-PSDB; AAZ43872.
XX
PT Ki-67 gene antisense oligonucleotide
XX
PS Disclosure; Page 19-29; 36pp; German.

Db f645 vda1sqf1nnyasvviid---dhpevtviedpqsn---lnddgftevsvkq----- 1691
 QY 1196 RKDVGEENGEORER-----NKNNDSDIDYAIRASISEATKRGKMNMLAVRIKD 1248
 Db 1692 qrlqdeerrkkeeqvivwnkknane-----kgrsqtksl----- 1727
 QY 1249 FLERIVKHGGIDLEWLRSPDPAKADYLLSIRGLKSVCEVRLTLHNLAFPPVDITNWG 1308
 Db 1728 -----pprfakkqatgiq---qasqasvvpplaspaplpstsaas 1763
 QY 1309 RIATVRMGWVP--LQPLPSLQHLLELIPVLESTQKFLWPRCLKDQRTLYELHQLITF 1366
 Db 1764 vpastaplattpvpastsa-----vpbas-----tl 1792
 QY 1367 GKVFTCKRPNCNACMRGECRHFASASARLALPAPERSLTSATIPVPESFP----- 1422
 Db 1793 apvlstap-vpaspi-----apvasasvsvpastsaaitssapasapapt 1844
 QY 1423 -----EVAIPMI---ELPLPLEKSLAS-GAPSNRENCEPIIEEPASPGOECEITEESD 1471
 Db 1845 ilasvstpasvtilasasipi-lasalastapt---apaasspaap-----vitapt 1894
 QY 1472 TEDAYNEDPEIPTIKLINIEQFGMTREHMERNELOEGDMSKALVALHHTTTSIPTPK 1531
 Db 1895 i-----pasapt-----asvplapasasapapa 1917
 QY 1532 LKNISRLRTEHOVYELPDSHRLDGMKREDDPSPYLLAIWTPGETANSAPPEKCGG 1591
 Db 1918 ptpvs-----apnpappap-----agtqaqthkvpqnlq----- 1947
 QY 1592 KASGKMCDFETCSECN-----SLREANSQTVRGTLIPCRMTAMRGSFPLNGTYFQVNELF 1646
 Db 1948 -----ttsqskqpppsiripsaqtpngt----- 1971
 QY 1647 ADHESLKPIDVPRDW-----IWD 1665
 Db 1972 -dyvasgksidtpqshgtltaelwd 1995

RESULT 3

AAR27204

ID AAR27204 standard; Protein; 1805 AA.

XX AC AAR27204;

XX DT 20-MAY-1998 (first entry)

XX DE Rat nestin.

XX KW Intermediate filament; central nervous system; brain tumour;

XX OS Rattus norvegicus.

XX PN W09214821-A.

XX PD 03-SEP-1992.

XX PF 21-FEB-1992; 92WO-US01375.

XX PR 22-FEB-1991; 91US-0660412.

XX PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX PI Lendahl U. McKay RDG;

XX DR WPI; 1992-316175/38.

XX DR N-PSDB; AAQ28398.

XX PT Diagnosis of pre-disposition to brain tumours - using DNA

XX PT encoding nestin which distinguishes neural multi-potential stem

XX PT cells from neuronal, glial and muscle cells

PS Claim 3; Fig 1; 63pp; English.
 XX The amino acid sequence of the nestin gene which was deduced from
 CC the nucleotide sequence, suggests that nestin is a member of the
 CC intermediate filament protein family. The rat nestin amino acid
 CC sequence shows 75% similarity with the human nestin sequence. There
 CC is more than 60% identity between the two sequences. Antibodies to
 CC nestin protein can be used in vivo diagnosis of brain tumours.
 CC See also AAQ28399 for the human nestin gene.
 XX SQ Sequence 1805 AA;

Query Match 2.3%; Score 210.5; DB 13; Length 1805;

Best Local Similarity 18.7%; Pred. No. 7.6e-07;

Matches 283; Conservative 216; Mismatches 564; Indels 447; Gaps 72;

QY 60 PRKPAELPKVVVEGPKRKRKAATOEKVKSK--ETGSAKKKLNKESATKKPANVGDMNS 117

Db 159 prrp-----papphripgapevedlarlrgewrgavrdygervahmesslgqare 210

QY 118 KSPVET--LKSCRKALNFDLENPGDARGDSESEIVQNSSGANSFSEIRDAI--GGTNGSF 174

Db 211 rlsqavrgarecr-----levqqiqadrsiqerrealqrlegw 251

QY 175 LDSYSQIDKTN--GLGAMNOPLEVSMGNOPDKLSTGAKLARDQOPDLLTRNOCCOPPVATQ 233

Db 252 qdrilatdkfqlavealeqekqqlsqiaqileggqqlah-----lkmnslslevaty 303

QY 234 NTQPFMENQAWLQKMLNOLIGFPFGNQOOPRMTIRNQOQCLAMNQOQPMYLIGTRPPALVS 293

Db 304 rtilaeasnr--lqtpgrgsqaslgfidpkl-----kpnf-lgipedqyl--gsvlpalsp 354

QY 294 GNOQLGGPOGNKRPI--FLNHOTCLPAGNQLYGSPTDMHQLVMSTGGQOHLKLNQOOPG 351

Db 355 tsfspipntletptvafiktqeflqartptlas--tpippiseap-----cpnn 402

QY 352 SLIRGQOPCVPLIDQOPATP---KGFTLNQWVATSMSSPG-----LR 391

Db 403 aevraqevplsalqtqapeplwlkatvypssaailpeleepgkqgghfddltlatnln 462

QY 392 PHOSQVPTTYLHVESVSRIILGTTCORSRAPAYDSLOODIHQGNKYI-----LSHEI 446

Db 463 phh-----pt-----leakdgessesrvasi--fgedegqlwelvekeadievkv 505

QY 447 SNGNGCKKALPONSSLPTPIMAKLEEARGS---KROYHRAMGOTEKHDLLAQIAQSQ 502

Db 506 enssaqk---tqesgldt-----eetqdsqgplqketikalg--eepmlsklqnyeta 554

QY 503 DVERHNST-----CYEYLDAAKTKIKVVOENL-----HGMP--PEVIEIDDDPTD 548

Db 555 gkencnssteghlgtlegpekeqkqplksleeknvesektengvpvlsellgkedtrte 614

QY 549 -----GARK-----GKNTASISKAGSKN-----SSPVKKT 574

Db 615 dqelmspgkgtlkrfsslgkesqevrvpskegnlswtafkeeshp1gfpgaedqm1erl 674

QY 575 AEKEKICVPTKPAKKG---RAGRKKSVPPPAH-----ASEIQLMQPTPK 616

Db 675 vekedqfprspeeedqacrp1qkenqep1gyeaaegq1lerliekesqeslsrpeeed 734

QY 617 TPLSRSKPKGK---GRKSIQDSGKARGPSGELLQCPDSIAEIIYRMQNLVIGDK--EREQE 671

Db 735 qeagrs1qkenqep1gyeaaedqm1er-----liekesqes1kspenqr1gkplerenq 789

QY 672 QNAMVLYKGDGALVPYSEKRRKPRPKVDIDDETRINLLMGKDEKEDG--EEKDKKKEK 730

Db 790 kslrlyleengetfvplesnrq1rslrseveeeedq1r1kplekvsqds1gslaeenvqplr 849

QY 731 WVEEE-----RRVFRGRADSF1ARMHLVQGRDRFRFPWKGVSVVDSVIGVFL-----TONVS 780

Db 850 yleedddcinkslledkthks1gs1edrngds1iip-----qesetqvs1rpeeedq1rv 904

```

QY 781 DHLSSAFMSLAARPPKLSRSDERNV--RSVVVEDEPGCILNLEINETSQEKVQHPSS 838
Db 905 nllkes-----qefsrseeeqvmerslegenhe-----slssvekedqmvve--s 949
QY 839 DMEVSVDSG-----SKEQLRDCNSGIERFNFLEKSIQNL-----EE 876
Db 950 qlekesqsgsklesdesqetfopleknaeslrslagqdeeqkleqetqqltravgneq 1009
QY 877 EYLSODSDP-----AFOSCGRGVSCS-CSKSDAEFPFTR-CETKTVSGTS-- 922
Db 1010 mavspekvdpelpkplgndqelarslgkendesvlslekgietvksleteilepleta 1069
QY 923 -----QSVQFGSPNLSDEICLQNERPHLYE-----GSGDVQKQETTNVAOKKPDLEKT 971
Db 1070 eedlerrksidteplwstevaretvepedpgslgsldenretltslekesqelsal 1129
QY 972 MNW-----KDSVCFQOPRNDTN-----WQTPSSSYEQ----- 999
Db 1130 gkwnvetrvdsqqclqveeglleqehqeslrevkqelpssgnqgrwvdevgkavgqea 1189
QY 1000 -CATRQPHVLDIEDFGMQGEG-----LGYSWM--SISPRVDRVKNKNVPR 1041
Db 1190 platgvgtedkaehlrqgggeeeaaegellqdivgeawslgssepkeqrvpaealdn 1249
QY 1042 RFRGGSVPREFGTQIIPSTHELPGMGLSGSSAVQEHQDD-TQHNQODEMKNASHLQ 1100
Db 1250 ---leggal-----evp---vadsmpevterdedraqagqedsievtlgile 1289
QY 1101 --KTFDLILNSSECLTQSSKQNTDCLPRDRTAEDVDPPLSNSSLOWILVESNSS 1158
Db 1290 aartglel-----eqevvgledprhfareaaipsigeesvk----- 1326
QY 1159 NKEQTAVEYKETNATILRMKGTGLADGKKPTSQWDSLRKDVEG-NEGROKNNKMSDID 1217
Db 1327 -----akiaggleg---pgkep-----keagaldsgilelpktsseale 1362
QY 1218 YEAIRRASISEIAKERGMNMLAVRIKFLERIVKDHGGIDLEWLRESPPDKAKD-- 1275
Db 1363 cqhgee---sesmegweeeas-----le--tsdhegsdapqpr--ppteedeeg 1405
QY 1276 YLLSIRGLGLKSEVCVRLTLHLNLAFFVDNVIAGVRMGWVPLOPLPESLQLHLELYP 1335
Db 1406 aqaaltapapkile-----pcspipiltadah--elqp 1435
QY 1336 VLESTQKFLW 1345
Db 1436 qaegiqeagw 1445

RESULT 4
AAR60126
ID AAR60126 standard; Protein; 1805 AA.
XX
AC AAR60126;
XX
DT 21-MAR-1995 (first entry)
XX
DE Rat nestin protein is useful to identify brain tumours.
XX
KW nestin gene; brain tumour; neoplastic cells; glial; neuronal;
KW muscle; neural multipotential stem cell; mammalian brain; detection;
KW diagnosis; medulloblastoma; glioblastoma; oligodendroglioma.
XX
OS Rattus rattus.
XX
FH Key Location/Qualifiers
FT Misc-difference 167..168
FT /note="encoded by CACGG"
FT Misc-difference 172..173
FT /note="encoded by CCG G GCC"
XX
PN US5338839-A.
XX

```

```

PD 16-AUG-1994.
XX
PF 12-APR-1988; 88US-0180548.
XX
PR 12-APR-1988; 88US-0180548.
PR 02-JUN-1988; 88US-0201762.
PR 25-OCT-1990; 90US-0603803.
PR 22-FEB-1991; 91US-0660412.
PR 19-MAR-1992; 92US-0853913.
XX
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Lendahl U, McKay RDG;
XX
DR WPI; 1994-263332/32.
DR N-PSDB; AAR70447.
XX
PT Nucleotide and protein sequences for human and rat nestin -
PT distinguishes neural multipotential stem cells and brain tumour
PT cells from more differentiated cell types; for use in the
PT diagnosis of brain tumours
XX
PS Claim 4; Column 35-44; 45pp; English.
XX
CC This sequence is the rat nestin protein encoded by AAR70447. Its
CC degree of sequence homology to intermediate filaments in the core
CC domain, its dissimilarity in the head and tail domains, and its
CC different splicing pattern suggest that it defines a new class of
CC intermediate filament protein. Nestin protein expression distinguishes
CC neural multipotential stem cells and brain tumour cells from the more
CC differentiated neural cell types (eg., neuronal, glial and muscle
CC cells of the adult brain). The nestin protein can be used in diagnosing
CC tumours of the brain, such as medulloblastomas, glioblastomas and
CC oligodendroglioma. (See also AAR60127).
XX
SQ Sequence 1805 AA;

```

```

Query Match 2.3%; Score 210.5; DB 15; Length 1805;
Best Local Similarity 18.7%; Pred. NO. 7.6e-07;
Matches 283; Conservative 216; Mismatches 564; Indels 447; Gaps 72;

```

```

QY 60 PRKPAELPKVVEGPKRKPRKAATQEKYKSK-ETGSAKKKNLKESATKKPANNVGDMSN 117
Db 159 prrp-----papphrpapgapevediarrigevrvagvdrdyqervahmesslgqare 210
QY 118 KSPVET--LKSCRKALNFDLENPGDARQDSESEIVQNSSGANSFSEIRDAI-GFTNGSF 174
Db 211 risqavrgarecr-----levqqlqadrdslqerrealqrlgrw 251
QY 175 LDSVSQIDKTN-GLGAMNQPLEVSMGNQDPDKLSTGAKLARDQDPDLLTRNQOCQFPVATO 233
Db 252 qdrlqatdkfqlavealeqekgqlsqlilegggqlah-----lksmlslevaty 303
QY 234 NTQFPNENQAWLQMKNLQIGFPFGNQOPRMTIRNQOCLAMGNQOPMYLIGTPPALVS 293
Db 304 rtlleaensr--lqtprgsgaslgfldpkl-----kpnf-lgipedgyl-gsvlpalsp 354
QY 294 GNOQLGGPGGNKRPI--FLNHQTCLPAGNLQYSGSTDMHQLVMSTGGQGHLLIRNQOPG 351
Db 355 tsfpplntletpvtafktqefiqartptlas-tpippiseap-----cpgn 402
QY 352 SIIRGQQPCVPLIDQOPATP---KGFTHLNQMVATSMSSPG-----LR 391
Db 403 aevraqevplslitqaapeplwlkatvpssailpelleepgkqgghfpdditslatnln 462
QY 392 PHSQSQVPTTYLHVESVSRILNGTTGTCCORSRAPVDSLQODIHQGNKYI-----LSHEI 446
Db 463 phh---pt-----leakdgesesrvssi--fgedegqiweiveheadlevkv 505
QY 447 SNGNGCKKALPONSSILPTIMAKLBEARGS---KROYHRAMGQTEKHDLNLAQQAQSQ 502
Db 506 ensaqk---tqesgldt-----eetqdsqplqketikalg--eepimskiqnyeta 554

```

Qy	503	DVERHNST	-----CVEYLDAAKTKIQKVOENL-----HGMP-----PEVIEIEDPTD	548
Db	555	gkencns	thgltigpekeqkqiplakeknvseksktlengvvpvisellgkurt	614
Qy	549	-----GARK	-----GKNTASISGASKGN-----SSPVKKT	574
Db	615	dqelmspgtkl	fsslgkesqevrvpskegnleswtafkesqhlpgfpaedqmlerl	674
Qy	575	AEKEKCI	VPKTPAKKG-----RAGRKSVPVPAH-----ASEIQLMQPTPPK	616
Db	675	vekedqsfp	rpspeedqearcpqlkenqeplyeaaegqilerliekesqeslrpspeed	734
Qy	617	TPLSRSPKPK	-----GRKSITDSGKARGSGELLCCQDSTAETIYRMQNLYLGDK-ERQDE	671
Db	735	qeagrs	lqkenqeplyeaaedqmler-----liekesqlkspeenqrkgkplerenq	789
Qy	672	QNAWVLY	KGDCALVPYESKRRKPRPKVIDDDTTRITWLLMGKGKEGD-EEDDKKKEK	730
Db	790	kslryle	enqetfvplesnqrplrsaleveeeeqrvpklevsqdslgslaeevqplr	845
Qy	731	WWEZE	-----RRVTRGRADSFIAHMLHVQODRRFPSPKGVSDSVIGVPL-----TONVS	780
Db	850	yleedd	cinksllektkhslgledrngdsliip-----qesetqvalrpspeedgriv	904
Qy	781	DHLSSSA	FAFMSLAARFPFKLSSSREDERNV-RSVVVEDEPGCILNLNIPSWQSKVOHPS	838
Db	905	nhlekes	-----qefsrseeeeqmrslegenhe-----slsvekedgmve--s	949
Qy	839	DMEVSGV	DVG-----SKELRDCNSGIERNFLEKSIONL-----EE	876
Db	950	qlekes	qdskslidesqetfpglekenaeslrslagdqeeqkleqetqqlravneq	1009
Qy	877	EVLSSQSD	FDP-----AIFQSGRVGSCS-CSKSDAEFFPTR-CETKTVSGTS--	922
Db	1010	mavsppek	vdpeipklpndqelarslgkendesivlskegiecvksletellepieta	1069
Qy	923	-----QSVQTS	GNPLSDEICLQNERPHLYE-----GSGDVQKQETTNVAKKPDLEKT	971
Db	1070	eedlerrks	idteplwstevaretveppedeppsgslgsdenretltslekesqelssl	1129
Qy	972	MNW	-----KDSVCFQOPRNDYN-----WOTTPSSSYEO-----	999
Db	1130	gkwnvetr	vedsqclqveegiqeehqeslrrevkqelpssgnqqrwedvvegkavgea	1189
Qy	1000	-CATRQPH	VLVIDFCMGQEG-----LGYSNM--SISPRDVRVKNKNVPR	1041
Db	1190	plattvgv	tedkaellhrqgggeaaaaagellqdivgeawslgssepkqrvpaeadln	1249
Qy	1042	RFRQGS	VPREFTGOTIPSTPHELPGMLGSSSAVQEHODD-TQHNOQDEMKNASHLQ	1100
Db	1250	---leggal	-----evp---vaqspmvterdedraqageqdsievtlgle	1289
Qy	1101	--KTFD	LLNSSECLTROSTTKONITDCCLPRDRAEDVDVPLSNSSLQNILVESNSS	1158
Db	1290	aartglet	-----eqevvgledprhfaraealppsigeesvk-----	1326
Qy	1159	NKEOTAVEY	KETNATLIREMKGTADGKKKPTSQWDSLRKDEVEG-NEGRQERNKNMDSID	1217
Db	1327	-----akia	qleg---pgkep-----keagaldsgillelpktsseale	1362
Qy	1218	YEAIRRAS	ITSEIAIKERGMNMLAVRIKDFLERIVKDXHGDIGLEWLRESPPKAXD--	1275
Db	1363	cqhhee	---sesmegweeeas-----le--tsdhegdsapqpr--ppeteedeg	1405
Qy	1276	YLLSIR	GLGLKSVCEVRLLTLHLNLAFFVDNVGRIARVMGWVLPQLPESQLHLLLYP	1335
Db	1406	aqaaltap	gpklle-----pscpipiltcda--elqp	1435
Qy	1336	VLESIO	KFLW	1345
Db	1436	qaegiqe	agw	1445

RESULT	5
AAV85565	
ID	AAV85565 standard; Protein; 2432 AA.
XX	
AC	AAV85565;
XX	
DT	07-JUL-2000 (first entry)
XX	
DE	Human homologue of UNC-53 (Hs-UNC-53/2) sequence.
XX	
KW	UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
KW	anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
KW	antisclerotic; antimetastatic; anti-arthritis; autoimmune disease.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	Region
FT	1..89
FT	/note= "this region can be replaced with one of the
FT	three sequences shown in AAV85566 to AAV85568;
FT	this creates three variants at the N-terminus"
FT	Misc-difference 1018
FT	/label= Asp or Glu
FT	1776..1778
FT	/note= "present or absent depending upon the allele
FT	from which the protein is translated"
XX	
PN	W09963080-A1.
XX	
PD	09-DEC-1999.
XX	
PF	02-JUN-1999; 99WO-EP03848.
XX	
PR	03-JUN-1998; 98GB-0011962.
XX	
PA	(JANC) JANSEN PHARM NV.
XX	
PI	Luyten WHML, De Raeymaeker MC, Geysen JJGH, Bogaert TAOE;
PI	Maerten LJS, Verhasselt P, Van De Craen M;
XX	
DR	WPI; 2000-116370/10.
DR	N-PSDB; AAA07836.
XX	
PT	Novel proteins and nucleic acids e.g. for treating neurodegeneration
XX	
PS	Claim 95; Fig 1d; 146pp; English.
XX	
CC	The invention provides vertebrate (human) protein homologue of a UNC-53
CC	protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
CC	their plus ends. The UNC-53 sequences are used to promote neural
CC	regeneration, revascularization and wound healing; also for treating
CC	neurodegenerative disease, acute traumatic injury, fibrotic disease and
CC	autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
CC	polynucleotides can be used for recombinant production of the proteins,
CC	as a source of probes for detecting allelic variants and polymorphisms,
CC	for sequencing genomic DNA and for detecting UNC-53 expression; and as
CC	source of therapeutic antisense sequences. Cells that express the
CC	protein are used to identify regulators of cell shape, growth, motility
CC	and migration. They can also be used to identify proteins that are
CC	involved in signal transduction pathways also involving UNC-53, and to
CC	identify compounds that alter attachment of UNC-53 to microtubules. A
CC	target gene coupled to a UNC-53 encoding sequence may be used to deliver
CC	the target gene to a cellular microtubule or its plus ends. The present
CC	sequence represents the amino acid sequence of the second human homologue
CC	of UNC-53, designated hs-UNC-53/2.
XX	
SQ	Sequence 2432 AA;

Query Match 2.3%; Score 208.5; DB 21; Length 2432;
Best Local Similarity 18.7%; Pred. No. 1.8e-06;
Matches 273; Conservative 171; Mismatches 547; Indels 469;

CC The polypeptides are useful for providing analogues of gravin in the
CC study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC interactions between gravin and kinase. The peptides are involved in the
CC modulation of gravin-kinase interactions.

XX SQ Sequence 1780 AA;

```
Query Match          2.1%; Score 195; DB 19; Length 1780;
Best Local Similarity 17.8%; Pred. No. 1.2e-05;
Matches 332; Conservative 262; Mismatches 690; Indels 584; Gaps 84;

QY 105 ATKPNVGMNKSPEVTLKCRKALNFDLEPGDARQDSEISEIVQNS-----SGAN 158
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 23 aepepgggpsaeapdt-----adpaa-asdpakllkngqlstingva 69

QY 159 SFSIRDAIGGTNGSFSDVSQDKTNGLGAMNQPLE-----VSMGNQPD 203
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 70 eqdelsiqegdlng-----qkalgngqalnsqeeevivtevgqrdsedvserdsd 122

QY 204 KLSTGAKLARDQOPDLLTRNQOC--QFPVATON-----TQFPWENQQAWLQKN--OLIGFP 256
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 123 ematkavvhdttdgqeenrnieqipsesnlleelcq-ptesqandigfkvfkfvfk 181

QY 257 FGNQQR-----MTIR-----NQQCLAMG-----NQQPMYLIGTPRPA 290
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 182 ftkvkkktekptvqlitvkdkdegaagagdhqpslgagaesekspkq--stekpe 239

QY 291 LVSNQOGLGPOGNKRPIFNHOTCLPAGNOLVGSPTDMHOLVSTGCGQHGLLIKNQOP 350
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 240 etlkreq-----shaelsp-----paesggaavee--keegseakeqep 276

QY 351 GSLIRGOQPCVPLIDQOPATPKGPTHLMQVATSMSSPGLRPHSOSOVPTTYLHVSVSR 410
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 277 sk--saesptspvtsetgskfkf--ftqwgagwrktsfrkpkedeveas----- 323

QY 411 ILNGTTGTCORSRAPYDLSQDDIHQGNKYLSHEISNGNGCKKALPQNSLFTPIMAKL 470
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 324 -----ekkkdeqekvdeed---gkaevaseklcaseqahpqaesaheprlsae 373

QY 471 EEAR-GSKROYHRAMGTEKHDLNLAQOIAQSQDVERHNS-----TCVEYLDAAKKT 522
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 374 ekvelpseeqvgsgpseeekaplatevf-dekievhqeevvaevhvstveerteekt 432

QY 523 KIQK-----VQVENLHGMPPEVETIE-----DDPTDGARKGKNTASISKGA 563
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 433 eveetagsvpaeeilvgmdaepqaepakelviketcvsgedptqgadlspdekvisk-p 491

QY 564 SKGNSSPVKKTAEKICIVPKTPAK-----KGRAGRKSVPPPAHASEIOLWQPTTP 615
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 492 pegvvsevemlssqermkvqgspklkiftstgklksgkkgkrggdeesgehtqtpa 551

QY 616 KTPLRSKPKGKGRKSLQDSKARGPSGELLG-QDSTAEIILYRMQNLVLDGKEREQONA 674
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 552 dspdsgeeqgessasspepe-----eltclcklaev-----qqdgaeeaga 595

QY 675 MVLYKGGG-----ALVPYES-----KKRPRP-KVDIDDETRINWLLMG----- 713
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 596 -----tsdgekkrregvtwafkkmvtpkkrvrpseedkedeldkvksatlsstestase 651

QY 714 -----KDEKEGDEKDKK-----EKWWE-----EER 736
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 652 mqeemkgsvveepkprkrdvtsvswaeallvcgsskkrrrrrrssdeegppkamgghq 711

QY 737 RVFGRADSTFARMHLVQGRDRRSPNKGSVVDSVIGVFLTQNVSDHLSS-SAFMSLAARF 795
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 712 kadeagkdketgtdgfiagsqehdpqggsspeqag---sptegegvstwestfkrlyt-- 766

QY 796 PPKLSSSREDERNVRVSV----- 813
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 767 prkksksleeksedsiaagsvehstpdtepgkeeswvskkflipgrkrkrpdkgkeqap 826

QY 814 VED--PBGCILNLEIPSWQEKVQHPSDMEVSGVDSGSKQLR-DCSNSIERFNFLEKS 870
```

```
Db 827 vedagtpga-----nedds-----dvpavvplseydarekmeaaggqgkgaep--qka 875
QY 871 IQNLEEVLSQDSFDPFAIFQSCGRVCSKSCSDAFPTTRCTETKTVSGTSQSVQIGSP 930
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 876 atevskelsesq-----vhmmaaavadgtraatifeerspswisavtpele 922

QY 931 NLSDEICLOGNE-----RPHLYEGSGD--VQKQETTNVAQKKPD--- 967
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 923 qveaeaalteevlereviaeeepptvteplpenreargdtvseaealtpeavtaetaag 982

QY 968 -----LEKTMNWKDSVCFQPRNDTNMQTTTPSSSYEQCATRQPHVLDIEDFGM 1015
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 983 plgseegeteasaeettemsvsqtldspdtteeatpvqve-----ggvptideer 1036

QY 1016 QGEGLYGYSWMSISPRVDRVKNKNVPRFRFGQGSVPREFTGQIIPSTPHELPGMGLSGSS 1075
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1037 rtqev-----lqavaevkvees-----qlpdtg--gpe 1062

QY 1076 SAVQEHQDDTQHNOQDEMKNKASHLOK----- 1101
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1063 dvlqpqv-raeaeerpeeqaeasglkktedvvllkvdaqaekteftgkvvgqttesfek 1121

QY 1102 --TFDLLNSESSECLTROSST-----KQNTDCLPRDRTAEDVVDPLSNNSLQNLV 1153
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1122 apqvtesiesseelvtctqaetlagvksqemvmeqalpdp--svetpdtsetdstpvadfd 1180

QY 1154 ESNSNKQETAVYEKETNATILREMKGTLADGKKPTSQWDSLRKDDVEGREGROERNK--N 1211
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1181 apdttkde-iveiheenevhlvprgtaeavpaqkerppapssvfqeetkeqskmed 1239

QY 1212 NMDSIDYE-AIRRASISEISEAIKERGMNNMLAVRKD--FLERIVKDHGGID--LEWLR 1266
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1240 tiehtkvevetvslsktegtqe--adqvadektdkvffegl---egsidgtitvr 1294

QY 1267 ESPDPKAKDYLLSIRIGLGLKSVCEVR--LTLNLNAPPVDTNVGR-IAYRWGVPLOPLP 1323
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1295 ekvte-----valkgegteeeackddalelqshakspvverenvvqverekteap 1348

QY 1324 ESLQLHLE--LYPVLESIQKFLWPLRLKLDQRTLYELHVQLITFGK----- 1368
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1349 thvneeklehetavtveevskql-----lqtnvnpildgakevsslegspccc 1397

QY 1369 ----VFCTKSRNCNACPMRGECRHFASAYASARLALPAPEERSL-TSATIPVPPESEFP 1422
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1398 lgqeeavctkiq-----vqseasftltaaeekvlgetenilletgetle 1443

QY 1423 PVAIPMIELPLPLEKSLASCAPSNRENCEPIIEEPASPGQEC-----TEITESIEDIAYN 1478
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1444 pagahlv-----lee---ksseknedfaahpgedavptgpcqakstpvivsattkkgls 1495

QY 1479 EDPEIPTIKLNTIEQFGMTLREHMERMELQEGDMSKALVALHPTT--TSIPTPKLKNIS 1536
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1496 sdlegekttslkws-----devdeqacvkvsvailedlepengllelektssklvq 1549

QY 1537 RL-----RTEHQVYEL-----PDSHRL-IDGMD-----KREDDP----- 1565
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1550 niigtavdqvrteetatemltsetlqtqahvfkadsqdaggetekegeepqasadepl 1609

QY 1566 -----SPYLLAIWTGGETANSQA-----PPQKCGGKA 1593
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1610 tsakeesestavqgahsdiskmseasektmtvevegstvndqgleevlpseeegggag 1669

QY 1594 SGKMCDFDE 1601
Db 1670 tksvpedd 1677
```

RESULT 7
AAB15380
ID AAB15380 standard; Protein; 1780 AA.
XX
AC AAB15380;

XX 26-JAN-2001 (first entry)
DT Human gravin protein sequence.
DE
XX
XX
KW Human; gravin; PKA RII binding site; myasthenia gravis;
KW kinase anchoring protein; CAMP dependent protein kinase.
XX
OS Homo sapiens.
XX
XX

XX Key Location/Qualifiers
FH Binding-site 265..556
FT /note="PKC binding site"
FT Binding-site 1526..1582
FT /note="PKA RII binding site"
FT Region 1537..1563
FT /note="PKA anchoring site"
XX
XX

XX US6090929-A.
XX
XX 18-JUL-2000.
XX
XX 19-DEC-1997; 97US-0994570.
XX
XX 19-DEC-1996; 96US-0769309.
XX

XX (UYOR-) UNIV OREGON HEALTH SCI.
XX

XX Klauck TM, Scott JD, Nauert JB;
XX

XX WPI; 2000-523763/47.
XX N-PSDB; AAA74903.
XX

XX Novel polynucleotides useful for detecting gravin in patients suffering
XX from Myasthenia gravis encodes CAMP-dependent protein kinase-binding
XX polypeptide and protein kinase C-binding polypeptide of gravin
XX

XX Claim 1; Column 35-45; 34pp; English.
XX

XX The present sequence is the protein sequence of human gravin. Gravin is
XX an A-kinase anchoring protein (AKAP) which is involved in the
XX localisation of CAMP dependent protein kinase A (PKA) via interactions
XX between the RII binding region and the PKA regulatory subunit RII. Gravin
XX is also an antigen found in myasthenia gravis sufferers, and it is
XX thought that antibodies to it may be useful in modulating the binding of
XX PKA, and thus aid in the treatment of the disease. The gravin coding
XX sequence was isolated by first screening a human umbilical vein
XX endothelial cell cDNA library with serum from a myasthenia gravis
XX patient, and then searching a human heart cDNA library for sequences
XX resembling the isolated sequence. This was done because the first
XX sequence obtained was shown to be shorter than the full length cDNA.
XX

XX Sequence 1780 AA;
XX

Query Match 2.1%; Score 195; DB 21; Length 1780;
Best Local Similarity 17.8%; Pred. No. 1.2e-05;
Matches 332; Conservative 262; Mismatches 690; Indels 584; Gaps 84;

OY 105 ATKRPANVGDMSKSPETLTKSCRKALNFDLENPCDARQGDSESEIVONS-----SCAN 158
DB 23 aepepgggpsaeapdtt-----adpaia-asdpatlklqngqlstingva 69
OY 159 SFSEIRDAIGTNGSFLDSVQIDKTNGLGAMNPLE-----VSMGNQPD 203
DB 70 eqdelisqgdlnq-----qkgalnggalnsqeevivrtevgqrdsedvsrdsdk 122
OY 204 KLSTGAKLARDQPDLLFRNQOC-QFPVATQN----TQPPMENQAWLQMKN--OLIGFP 256
DB 123 ematksavvhditdggdeenrnieqipsesnleeltq-ptesqandigfkkvfkfvgfk 181
OY 257 FGNQOQPR-----WTIR-----NOOPCLAMG-----NOOPMYLIGTPRPA 290
DB

Db 182 ftvkkdktekpdvtvqltlcvkkdegeagaagdhqpslgageaasesepkq--stekpe 239
OY 291 LVSGNQQLGGPGGNKRPIFLNHTQCLPAGNQLYGSPTDMHQLVMSYTGQQHGLLKNQOP 350
DB 240 etllkreg-----shaeisp-----paesgqaveec--kegeekqekpe 276
OY 351 GSLIRGQOPCVPLIDQOPATPKGFTHLNQMVATSMSSPCLRPHSQSQVPTTYLHVESVR 410
DB 277 sk--saesptspvtsetgstfkf--ftggwagvrktsfrkpkedeveas----- 323
OY 411 ILNGTGTGQRSRAPAYDSLOQDIHQNKYILSHISNGCKKALPONSLPTPIMAKL 470
DB 324 -----ekkkeqekvdted---gkaevaseklaseahqpepaesaheprisaey 373
OY 471 BEAR-GSKRQYHRAMGOTEKHDNLNLAQIAQSDQVVERHNS-----TCVEYLDAAKKT 522
DB 374 ekvelpseeqvgsggpseeekpaplavevf-dekleivqeevvaevhvstveerteckt 432
OY 523 KIQK----VVOENLHGMPPEVIEIE-----DDPTDGARKGKNTASISKA 563
DB 433 eveetagsvpaeeivgmdeaepqaeapakelvkltketcvsgedptggadlspdekvlsk-p 491
OY 564 SKGNSSPVKKTAEKEKCIVPKTPAK-----KGRAGRKKSVPPPAHASEITQLWQPTPP 615
DB 492 pegvvsevemlssqermkvqgsplklftstgklklsgkkgkgrggdeesgehtqvpa 551
OY 616 KTPLSRSPKPKGRKSIQDSGKARPGSGELLC-QDSIAEIIYRMQNLVYLGDKERQEQNA 674
DB 552 dspdsqeeqkgesaasspeepe-----eitclekglav-----qgdgaeega 595
OY 675 MVLKGDG-----ALVPYIES-----KKRPRP-KVIDDDETTIRIWNLMG----- 713
DB 596 ----tsdgekkgregvtpwasfkvmtpkkrvrpsdeskedelkvksatissstestase 651
OY 714 ----KGDEKGDDEKDKK---EKWWE-----EKWWE-----EER 736
DB 652 mqeemkgsveepkpeepkrkvdtsvswaalicvsgskkrarrirssdeegpkmangdhq 711
OY 737 RVFRGRADSFITARMHLVQDRRFSPWKSVDVSVIGVFLVTQNVSDHLS-SAFMSLAARF 795
DB 712 kadeagkdketgdgilagsqehdpggsssqeqag---sptegegvstwesfkrllvt-- 766
OY 796 PPKLSSSREDERNRSVV-----
DB 767 prkkskleekeksedsiaagsvvehtpdpkeesvswikfliprrkkrpdgkqeqap 826
OY 814 VED--PEGCILNLNIPSWQEKVQHPDMEVSGVDSGSGKEOLR-DCSNSGIEREFNLEKS 870
DB 827 vedagptga---nedds-----dvpavvipseydarekmeaqaqkgaepqe--gka 875
OY 871 IQNLEEEVLSQDSFDPAPAFQSCGRVSCSCSKSDAEFPTRTCETKTVSGTSQSQTGSP 930
DB 876 atevskelisesq-----vhumaaavaadgtraatiieerspswisasvteple 922
OY 931 NLSDEICLOGNE-----RPHLYEGSGD--VQKQETNNVAQKRPD--- 967
DB 923 qveaeeallteeviereviaeeepvtvteipenreargdtvsvseaeitpeavtaetaetg 982
OY 968 -----LEKTMNWKDSVCFGQPRNDTNTQTPSSSVYEQCATRQPHVLDIEDFGM 1015
DB 983 plgseegtcaaaeeettemvsavsgltdspdtteatvpqve-----ggvpdleeqer 1036
OY 1016 QGEGIGYSWMSITSPRDRVKNKNVPRRFRQGGSVPRFTGQIIPSTPHELPGLMGLSGSS 1075
DB 1037 rtqev-----lqvaeekvees
OY 1076 SAVQEHODDTOHNOODEMNKASHLOK-----
DB 1063 dvlqpvq-raaeerpeeaeasgikketdvvlkvdqaektepftqgkvvgqgttpefsek 1121
OY 1102 --TFDLNLSSECLTQROST-----KQNTDGCCLPRDRTAEDVVDPLSNSSILQNLIV 1153
DB 1122 apqvtesiesseelvttcqaetlagvksqemymeapdpd-svetptdsetdgstpvadfd 1180

Db 1020 qltdevthtqskvqteeqyl-----emkmbddlifekyirnkseadlllremen----l 1070
Qy 572 KKTAKKEKCIIVKTPAKKGRAGRKSVPPPAHASFIQLWQTPPKTPILSRKPKGKGRKS 631
Db 1071 kgtmesvevkiadtkeleetiirkeqilhekkkyffqamqtiftpitplsdsippsk---- 1126
Qy 632 IQDSKARGPSGELLCQDSIAFIIVRMQNLVLDKER-----EFOENAMVLYKGDGAL 684
Db 1127 -----lvgegs-----qdpie--indyhnlialaternnmvcleternsl----- 1165
Qy 685 VPYESKKRKPVPKVIDDDETRIMWLLMGKDEKGEDEKDKKEKWEERRRVFRGRAD 744
Db 1166 -----keqvindntqlsl-----qagsieksdlqpkqglegev----- 1201
Qy 745 SFIARMHLVQDRRPSPKGVSVDSVIGVFLTQNVSDHLSSAFMSLAARPPKPLSSSRE 804
Db 1202 klllemell-----kghltdsqsliekqlenlevte-----klqtlqe 1240
Qy 805 DERNV-----RSVVVDDEPGCILNLEIPSWQEKVQHPSPDMEVS 843
Db 1241 emknitiernelqtnfedlkaehdskldlsenleqsietsqdelraaqeelreqkql--- 1297
Qy 844 GVDGSKQELRDCNSGIERENFLFKSITONLEEEVLSSQDSFDPAPAFOSCGRVGSCSK 903
Db 1298 -vds-frqqlldcs--vgisspn--hdavangeksvlgvnsalqsemllgerdelqtscka 1352
Qy 904 SDAEPTTTCETKTVSGTSQSVQTSQSPNLSDICLQGNR-----PHLYEGSGDVQKQ 956
Db 1353 lvselellrahvksvegenleittkklnglekeilgkseesevklsmenlknednklkeq 1412
Qy 957 -----ETTNVAQKKPD-----LKTWNKDSVCFG--QPRN----- 985
Db 1413 aeeysskenqfsleevsgsqkldvleivlkaqlkaaeerleikdrdyfelvqtantnlv 1472
Qy 986 DTNMQTTSSSYEQ-CATROPHVLIDIEFGQGGELGYSWMSISPRVDRVKNKNVPRFF 1044
Db 1473 egkietqladheedsidrrseemeikvlgeklerngyllrgekleisnkleilqke 1532
Qy 1045 RQGSVPREFTGQIIPSTPHELPGMLSGSSSAVOEHODDTQ---HNOQDQDMNKASHL-- 1099
Db 1533 metsvllkddllqklesllse--nllkenidttiklshsdtgaqlqktqgelqlaknlai 1590
Qy 1100 -----QKTFLLNLSSE--CLTQSSSTKQNTDGLPRDRTAEDV--VDPLSN 1144
Db 1591 aasdnclptqeketsadcvhpleekillllteelhqktneqekillhekeleaqvelkce 1650
Qy 1145 NSSQLNILEVSNSS-----NKEQTAVEYKETAFTILREMKGTADGKKPTSDWDSL 1195
Db 1651 vehlmskiesksleslqhekhdtqegllalkqmqvvtqekkelqgthenltaevdhl 1710
Qy 1196 RKDVE-----GNEGRQERN-----NNMDSIDYEAIR-RASITSEISEAIKERM----- 1238
Db 1711 kenieIglnfneaqkktkeqcllnnenkeleqshrlqceielelmslkkkesaletlk 1770
Qy 1239 -NNMLAVRIKDFELRIVKDHGIDLEWLES-----PPDKAKDYLLSIRGLGLKSVCEV 1291
Db 1771 eseqkvlnlndqemnmv-----lemeelknsgrtviaerdqlqddl-----resve-- 1816
Qy 1292 RLLTLHLNLAFFVDNMGVRIARMGVVPLQPLPESIQHLLELYPVLESIOKFLNPLCKL 1351
Db 1817 -----msiet---qddllrkaqaelqqgkdkvq-----eltsqisvlqe-----kisl 1856
Qy 1352 DQRTLYELHYQLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLALPAPERSLTS 1411
Db 1857 enqml---ynvatvketlserrddlnsqkhlfsietlslslkekefaleqaekdkada 1912
Qy 1412 A--TIPVPPEFPVAPIMELPLPLEKSLGASFPNRECEPIIEEPASQGCETEITE 1469
Db 1913 arktidi--tekisnleeqllqqatnlketl-----yeresliqckeglalntehretlk 1966
Qy 1470 S-DIEDAYNEDDPEIPIKLNIEQFGNLTREHMERNM-ELQEGDMSKALVALHPPTTISI 1527
Db 1967 skdialgkmederdeaankvialtekmssleeqnenvttlkegegek----- 2014

Qy 1528 PTPKUKNISRLRTERHQVYVELDPDSHRLLD 1555
Db 2015 etfyiqrpskqgssqmeelreslktkd 2042
RESULT 9
AAR41043
ID AAR41043 standard; protein; 1786 AA.
XX AAR41043;
AC AAR41043;
DT 22-MAR-1994 (first entry)
XX CD4-EBAL175 fusion protein.
DE Merozoite; Erythrocyte Binding Antigen 175; malaria; HIV; env;
KW human immunodeficiency virus; envelope glycoprotein; hybrid protein;
KW red blood cell; erythrocyte; AIDS; molecular machine.
XX Chimeric Homo sapiens.
OS Chimeric Plasmodium falciparum.
XX Key Location/Qualifiers
FT Region 1..371
FT /note= "residues 1-371 of CD4"
FT Region 372..1786
XX /note= "residues 20-1435 of EBA-175"
PN WO9318160-A.
XX 16-SEP-1993.
XX 10-MAR-1993; 93WO-GB00505.
XX 11-MAR-1992; 92GB-0005276.
PR 08-JUL-1992; 92GB-0014481.
PR 24-JUL-1992; 92GB-0015829.
PR 16-SEP-1992; 92GB-0019562.
PR 03-MAR-1993; 93GB-0004311.
PA (PREN/) PRENDERGAST K F.
XX Prendergast KF;
XX WPI; 1993-303474/38.
DR Anti-viral fusion peptide(s) - comprise viral-binding component
PT and malaria merozoite red cell binding component, for treating
PT e.g. HIV, and hepatitis
XX Claim 9; Page 44-47; 69pp; English.
XX The hybrid protein NH2-CD4(1-371)-EBA175(20-1435)-COOH is a
CC specifically claimed example of a fusion protein of the invention;
CC it comprises at least part of the CD4 molecule fused to a peptide
CC from a malarial parasite merozoite protein with affinity for red
CC blood cells. The fusion protein can bind free HIV in the blood to
CC red blood cells and consequently reduce viral titre, prevent
CC transmission of the virus and improve safety of blood transfusions.
XX Sequence 1786 AA;
SQ

Query Match 2.18; Score 191.5; DB 14; Length 1786;
Best Local Similarity 18.9%; Pred. No. 2.4e-05;
Matches 180; Conservative 141; Mismatches 340; Indels 289; Gaps 42;
Qy 441 ILSHEISNGNCKKALPSSSLPTPIMAK-----LEARGSKROYHRAHQTEKHDLNL 494
Db 651 mlsehkninnckn-ipqeelqitqwikewhgefllernrskipkscknt----- 702
Qy 495 AQQIAQSODVERHNSSTCVEYLD-----AAKTKTKQVVOENLHGMPPEVIEIE 543

```

703  -----lyeacekeidpcmkrydlwlrskfwhltlskeyetqkvpknaeny---likis 754
544  DDPTDGARK-----GKNTASISKASGNSSPVKTAKEKCIIVPKTPAK 588
755  enkndakvlllncdaeyskycackttltlvkvlngndntik---ekrehldldfsk 811
589  KGRAGRKSVPPPAHASEIQLWQPTTPKTPLSRKPKGKRSIQDSGKARGPSGELLQ 648
812  fg--cdknsvd-----tntkwecknpyi-----lstkd-----vc- 840
649  DSIAEIIYRMQNLVIGKERQEQONAMVLYKGDGALVPYESK--KRKPRPKVD-----ID 701
841  -----vpprrqelcgnidriyknllmikehilaiaiaiesrllrkyknkddkevckii 895
702  DET-----TRVWN-----LLMGKDEKEGDEKKKEK-----WWEERR----- 737
896  nktfadlrldi9gdyndlnsrklvknintnskyvhrnkndklfdewkwvjkdwvn 955
738  ----VFRGRADSFIAHMLVQGRDRFPWKGVSVDVIGVFLTQNVSDHLSSSAFMSLAA 793
956  viswvfk---dktyckeddiienipqfrfwsewgddy----- 989
794  RFPKLSSSRDERNRSVVVE-----DPEGCLINLETPSWQEKVQHPSDMEVSGVDS 847
990  -----cqdktmietlkeckepcedndckscnsykwiskkk----- 1029
848  GSKBQLRDCNSGIERNFLEKSTONLEE-----EVLSSODSFDPAIF-----QSCGRVGS 898
1030  -----eeynkqakqyqyqgkgnnykmysefiksikevylkkysekcslnf 1075
899  CSCSK-----SDAEPPTTRC-ETKTV-----SGTSQSVQTSQPNLSDFICLQGNRPHLYE 948
1076  edefkeelhsdyknkctmcpvkdvpislrnneqtsqeaeppeenteila-hrtetpsise 1134
949  G-SGDVQKQ-----ETTNVAQKKPDLEKTMWNKWDVCFGQPR 984
1135  gpkneqkerdddlslksisvpsensrpetdaktdtsnlklkgdvdisn---pkavigssp 1191
985  NDTNWQTPSSSYEQCATRQPHVLDIEDFGMGEGGLGYSMWSISPRVDRVKNKNVPRRFF 1044
1192  ndn-----invteqgdnisgvnskplsdvrdpkdeled----- 1225
1045  RQGSVPREFTGQIIPSTPHELPGMGLSGSSAVQEHDDTQHNOQDEMKNKASHLQKTF 1104
1226  -qnseseetvnhiskspsinngdsgsgsatvsessntglisiddrn-----gdtfv 1279
1105  ---DLLNSSECLTRQSTKQNTIDGC-LPRDRTAEDVDP-----LSNN---SSLQNLV 1153
1280  rtqdtan-tedvirkenadkdedekadeerhstseslsspeekmltdneggnslnheev 1338
1154  ESNSNKEQTAVEYKETNATILREMKGTLDAGKKPTSQWDSLRKDVEGNEGRQERNKNM 1213
1339  kehtsndnvgqsggvmnmvekelkdtl---enpssld-----eg-kaheelsepn 1388
1214  DSIDYEAIRRASISEIAKERGMNMLAVRIKFLERIVKHGGIDLE 1263
1389  ssdgdmsntppldntseetterisnneykvneredertitkeyedivlk 1438

```

RESULT 10

AAP81187
ID AAP81187 standard; protein; 1284 AA.

XX AAP81187;

XX 04-DEC-1990 (first entry)

XX Sequence encoded by a gene fragment coding for major protein of an A-type
DE inclusion body and a promoter region in poxvirus.

XX Vaccinia virus; vaccine.

XX

```

OS Cowpox virus.
FH Key Location/Qualifiers
FT Region 1..20
XX /note="Encoded by promoter region"
PN EP261925-A.
PD 30-MAR-1988.
XX
XX 22-SEP-1987; 87EP-0308372.
XX 09-SEP-1987; 87JP-0223972.
XX 02-SEP-1986; 86JP-0222194.
XX (TOFU ) TOA NENRYO KOGYO KK.
XX
XX Shida H, Funahashi S;
PI WPI; 1988-086185/13.
XX N-PSDB; AAN81538.
XX
XX Gene fragment coding A-type inclusion body in pox virus -
XX used in the construction of recombinant vaccinia viruses for use
XX as vaccines
XX
XX Disclosure; 3-10; 24pp; English.
XX
XX The gene is nonessential for proliferation of poxvirus and is homologous
XX with a corresp. gene of a vaccinia virus and can therefore be used for
XX construction of a recombinant vaccinia virus for use as vaccines. The
XX promoter present upstream of this gene is very strong and is adequate as
XX promoter for expression of an exogenous antigen cell.
XX
XX Sequence 1284 AA;
SQ

```

Query Match 2.1%; Score 190; DB 9; Length 1284;
Best Local Similarity 18.4%; Pred. No. 1.8e-05;
Matches 223; Conservative 159; Mismatches 387; Indels 442; Gaps 56;

```

QY 469 KLEEARSKQYHRAMQTEKH-----DLNLAQQIAQSQDVERHNSSTCTVEYLDAAKTKI 524
DB 414 kceecnggeykteleearkklitelnsdkiskrtier-----dsvykt-- 461
QY 525 QKVVOENLHGMPPEVIEDPTDGARKKNTASISGASKGNSPVKTAKEKCIIVPK 584
DB 462 -----eridrltkeikeld-----ingtdgdsdsdeidkktirelsdrer----- 506
QY 585 TPAKKGRAKRKSVPPPAHASEIQLWQPTTPKTPLSRKPKGKRSIQDSGKARGPSGE 644
DB 507 -----emrteleere-----ldtird-gkvegs--- 527
QY 645 LLCQDSTAEIIRYMONVLYLGDKEREQONAMVLYKGDGALVPYESKRRKPPKVDIDDET 704
DB 528 --cq-----relelsrml-----kdrdddlraeidkrr 554
QY 705 TRIWNLLMGKDEKEG---EEKDKKE-----KWMEERRVFRGRADSF-IARMH 751
DB 555 nvewelsrlrldikeckykedldkaktisnyvristleseiakyqqdrdtlsvrire 614
QY 752 LVQGDRRFPWKGVSVDVIGVFLTQNVSDHLSAPMSLAARPPKLSRSDERDNRVS 811
DB 615 leeertrvrdlesrlde-----ctrnqetqevdalrsrlrelenkladomesgg--- 665
QY 812 VVVEDPEGCLINLETPSWQEKVQHPSDMEVSGVDSKESQRLDCNSGIERNFLEKSI 871
DB 666 -----nltelsrlgskl---sdle-----rqlrecrgnate-isrlqyrl 701
QY 872 QNLEEEVLSSQDSFDPAIFQSCGRVSGCSKSDAEPPTTRCETKTVSGT"SQSVQTSQPN 931
DB 702 tdlterql-----ndcrrnne---nnadteremqrlrdritclerq----- 738

```

```
Qy 932 LSEDTCLQNERPHLYEGSGDVQKQETTNVAOKKPDLEKTMNWKDSVCFGQPRNDTNMOT 991
Db 739 Lsd-----crrnn-----esnadmer-emgrldrindldrqln----- 771
Qy 992 TSSSEYECATQPHVLDIEDFGMGEGGLGYSMWSISPRVDRVKNKNVPRFFROGGVP 1051
Db 772 -----ec-----krdngtcs-----seevnflktri----- 792
Qy 1052 REFTQGIIPSTPHELPGMGLSGSSSAVOEHODDTQHNQOQDENMKASHLQKTFDLNLSNE 1111
Db 793 -----rdlersleicskdeselysakykselgrar-----eqisnlq 828
Qy 1112 ECLTRQSTKQNTQICLPDRRTAEDVDVPLSNNSLQNLIVESNSS----- 1158
Db 829 eslrreresdk--tdsyyrreltre-----rnkiveleleknkcfctdnhakyi 874
Qy 1159 ---NKEOTAVEYKETNATILRMKCTFLADGKKPTQOWDSLAKDV---EGNEG----- 1204
Db 875 deinskktrisdlerqlaacsngsgndmdgdykreieslkrrelaecrrgnngshsdcky 934
Qy 1205 -----RQE--RNKNNMDSIDYAIRRASISEISEAIKERGMNMLA--VRIKDFLERIVK 1255
Db 935 ydeearveevkrlrqeltql-hedlkrateskndsykrelgrakvievekeleryfd 993
Qy 1256 D-----HGGIDLEWLRESPPDKADYLLSIR--GLGLSKSVECVRLTLHLNLAFFVDT 1305
Db 994 dsrlaackrhgd---emlr-----kiadlekklrdgngngngctsscefer----- 1038
Qy 1306 NVGRTAVRMGWVPLQPLPESLQHLLELYPVLESTQKFL-WPRL--C--KLDO----- 1353
Db 1039 --krlav-----leevrkmetlkslekfmeofrlqdcadklidrekerim 1083
Qy 1354 RFLYELHYQLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLALPAPEE----- 1406
Db 1084 kaerdlerei-----arkncggnpcerelesersnvkrleyqldaekvfykr 1133
Qy 1407 -----RLTSATIPVPPESPFPVPAIPMIEPLPLEKSLASGAPSNRENCEPIIE 1456
Db 1134 elerdrylssryltssdp-dekplpnytfpriev-----epittedtepkpve 1181
Qy 1457 PASPOQECTEITESDIEDAYYNEDPDEPTIKLNTIEQFGMTLREHMERNMELQEGDMSKA 1516
Db 1182 vvpssdvtepiisgy-----tpsvdaepe-----hpdlsevyqtsvs 1218
Qy 1517 LVALHTTTSIPTPKLNISRLRTEHQVYELPDSHR-----LLDGMKREPDDEPSP 1567
Db 1219 qvavtpp-----pkpetqiseyqdyselysasnntsknvsfelayldldkl--ddide 1272
Qy 1568 YLLAIWTPTGET 1578
Db 1273 yllnnimpekt 1283
```

RESULT 11

AA12454
ID AA12454 standard; Protein; 2453 AA.

AC AA12454;

XX 24-OCT-2000 (first entry)

XX HNRCR protein sequence.

XX Human; HNRCR; nuclear receptor coreceptor.

XX Unidentified.

XX CN1250094-A.

XX 12-APR-2000.

XX 06-OCT-1998; 98CN-0120919.

XX

```
PR 06-OCT-1998; 98CN-0120919.
XX (XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.
XX PI Yu L, Tu Q, Zhao Y;
XX WPI; 2000-400830/35.
DR N-PSDB; AAA60630.
XX Preparation of new human kernon acceptor co-repressor coding series and
PT the polypeptide.
XX Example 2; Fig 2; 58pp; Chinese.
PS The present invention describes a human homologue of nuclear receptor
XX coreceptor (HNRCR). The present sequence represents an HNRCR protein
CC sequence used in comparison with the human HNRCR.
XX Sequence 2453 AA;
```

Query Match 2.1%; Score 189; DB 21; Length 2453;

Best Local Similarity 17.8%; Pred. No. 6.4e-05;

Matches 287; Conservative 204; Mismatches 553; Indels 568; Gaps 67;

Qy 2 QSIMDSSAVNATEATEQND-----GSRQDVLEFDL-----NKTPOQKPSKRX- 43

Db 735 esapspsveaakssedssenaasrgntepvaeleattdpacaaspssavpttkpaeres 794

Qy 44 -----RKFPKVV-----VEGPKRKKPKPAELPKVVEGPKRKRKAATQE 86

Db 795 veaqvtdsaetaepmdvdheecgaegssvldppaptkadsvdpe-----mqvpe 845

Qy 87 KVKSETGSAKKNLKESATKPPANVGDSMNKSPVTLKSCRKALNFLENPGDARQDS 146

Db 846 ntashgedakderldestekteardv-----vvaqlerp-epgsodd 890

Qy 147 ESEIVONSSGANSFSE-----IRDAIGTNGSFLDSVSQIDKTNGL-----GAM 190

Db 891 ssatscdegvdgepergrvfpmdakpsalltpggsil--isspikpnllldlpqlghraav 948

Qy 191 NQPLEVSGNQPDKLSTGAKLA-----RDQOPDLTLTRNQCCQ 227

Db 949 lppm---vscpcnlpigtvpvsgyalyqrhlkamesalleepqrqeqvdlcrrsts- 1004

Qy 228 FPAVATQNTQFPMENQQAQLQMKNLII-GFPFGNQOP-----RMTIRNQOP 271

Db 1005 pcstks--pnrewevlqphqvitalpegvrlpttrppplipsskttvasekp 1061

Qy 272 CLAMGNQPMYLIQTPRPALVSGNQOLGGPQGNKRPFLNHOTCLPAGNQLYGSPTDMHQ 331

Db 1062 sfimggsisq---gtpgtylsshnday--pgeapkp-----svgsisiglprrqges 1107

Qy 332 L-----VMSGGOQHGLLTKNQOPGSLIRGQOPCVPLLDQOPATPQGFTHL 377

Db 1108 tkaapltyikgeefsrqsnsqpegllvraqheg--vvrtagav----qegsitrg-tpa 1161

Qy 378 NQVATSMSS-----PGLRPHSQSQVPTTYL----- 403

Db 1162 skisvetisslrsgsitqgtpalpqagiptealvqpvvrmpleesspekvreeaaskghv 1221

Qy 404 -----HVESVSRILNGTTCQCRSRAP-----AYDSLQODIHQG----- 437

Db 1222 iyegkshilsydniknaregt---rsprtahemslkrseavegskqgmmsrespsv 1277

Qy 438 -----NKYILSHSISNG-----NGCKKALPONSSLPPI 466

Db 1278 apleglralprgshdsldkervtvisgsmggtprataesfedgk--ypqkrespp 1335

Qy 467 MAKLEEARSGSRQYH-----RAMQTEKHDLNLAAQIAQSODVERH-----N 508

Db 1336 irafegaitkpkpydittikemgrsihelpqrdi-ltgesrktpevvqgstprliegsls 1394

```
Qy 509 SSTCEVYLDAAKTKIQKVQVQENLHG---MPPEVIEIEDDDTGDGARKG----- 554
Dy 1395 ggtplkfdnsggaikhnvksllitgspklprgmleivpenikvvekyedvkagepvr 1454
Qy 555 --NTASISKG-----ASKGNSSP-----VKTAKEKCIYPKT--- 585
Dy 1455 arhtsvsvsgsvlrstltheapkaqlspglyddasarrtpvsyqntisrgspmmnrtsdv 1514
Qy 586 PAKGRAGRKKSPPPAHASIQLWQTPP-----PKTPLSRSKPKGGRKS----- 631
Dy 1515 sssksasherkstltptqresipakspvgvdpivshpsfpdphrrssaagevyrshlpth 1574
Qy 632 ---IQDSCKARGSGELLCOQSDIAEI-----IYRMQN---LYLGDKEREQEONAMVL 677
Dy 1575 ldpampfhraldpaaayllqrlsptpgypsqyqlyamentrtqilndyitsqmqvnlr 1634
Qy 678 YKGDGALVPYESKKRPPK---VDID-----DETTRI----- 707
Dy 1635 pdvtgrispreqplglpypatrglidltnmpptilvphaggtstppmdrityipgtqvf 1694
Qy 708 -----WNLLMGKDEKDEKDKKKKWEERRVFRGRADSFIAARMHL 752
Dy 1695 pprpynaaslsphgpthlaaasaererererererereriaaapadlyl 1754
Qy 753 VQG-----DRFSPMKSGSVVDSVI----- 771
Dy 1755 rpgseqpgrgshgyvrpspsvrtqetlqrpstvfgtngtstvltpldptaqlrimpl 1814
Qy 772 ---GVFLTQNVSDHLSSAFMSLAARPPPKLSSSRED-----ERNVRS--- 811
Dy 1815 psggsisqgipasyryntaadaalydaaasapqmdvsktkeakheaaarleenlrssa 1874
Qy 812 VVDEPE-----GCILNLNEIPSWQEKVQHPDMEVSGVDSGS-----KEQ 852
Dy 1875 avseqqlleqknlekrsvctvcstsaipsgkqphasvvyseagkdkgppkksryeee 1934
Qy 853 LRDCSNGIERNFLEKSIQNLLEEIVSSQDSFDPFQSCGRVSGSCSKSDAEFFTR 912
Dy 1935 lrtgkttitaanfivli---trqiasdkda-----rergsgsdssslsshr 1981
Qy 913 CETKVTSGTSOSVOTGPNLSD-----EICLOGNERPHLYEGS-GDVOKQ 956
Dy 1982 yet-----asdaievispaesappgkqayqdpdmvkanqaenestrqyegpllhvrsq 2036
Qy 957 ETTNVAOKKPD-----EKTWNWKSVC-----FGOPR-----NDTNWQ 990
Dy 2037 qespsqgqplpsssegmgqvrthritladhicqitqdfarnqvpssqastcfq 2096
Qy 991 TTPSS-SYEQCATROPHVLDIEDFGMOGEGIGYSWMSISPRV-DRVKNKV--PRRFRQ 1046
Dy 2097 tpsalsstpvrtktsryspe-----sqstvlhprpgprvpenlvdksgsrp 2147
Qy 1047 GGSVPRFTGQIIPSTHELPGMGLSGSSSAVOBQDDTOHQNDENKASHLOKTFDL 1106
Dy 2148 gkspersh---ipsepye-----pisppqgpavhekqds-----lll 2182
Qy 1107 LNSSECLTROSKQONITDGLPRDRTAEDVVDPLSNSSLONLIVESNSS 1158
Dy 2183 sqryvdaeqrsdrspgsisylpsfftkiestspmv-kskkqefrkluss 2233

RESULT 12
AAR70232
ID AAR70232 standard; Protein; 1435 AA.
XX
AC AAR70232;
XX
XX
DT 22-SEP-1995 (first entry)
DE
DE P. falciparum SABP.
XX
KW SABP; stalic acid binding protein; binding domain; merozoite;
KW malaria; therapy; vaccine.
```

```
XX Plasmodium falciparum.
OS
XX
FH Key Location/Qualifiers
FT Domain 1..616
FT /label= Binding_domain
XX
XX WO9507353-A.
XX
XX 16-MAR-1995.
XX
XX 07-SEP-1994; 94WO-US10230.
XX
XX 10-SEP-1993; 93US-O119677.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
XX Wellens TE;
XX
XX WPI; 1995-123427/16.
XX
XX N-PSDB; AAQ83525.
XX
XX New erythrocyte binding domain polypeptide(s) - isolated from
XX Plasmodium binding proteins, used in diagnosis, treatment and
XX prevention of malaria
XX
XX Disclosure; Page 39-41; 81pp; English.
XX
XX Sequences from the SABP gene (given in AAQ83525) were PCR amplified,
XX expressed on the surface of COS cells and tested for erythrocyte
XX binding to identify the binding domain polypeptide. A pref. SABP
XX binding domain comprises residues 1 to about 616 of the SABP protein
XX (AAR70232). Recombinant binding domain was expressed in E. coli,
XX yeast, mammalian, insect, and in vaccinia virus and adenovirus-
XX infected cells. It provides protection against P. falciparum.
XX
XX Sequence 1435 AA;

Query Match 2.1%; Score 188.5; DB 16; Length 1435;
Best Local Similarity 18.3%; Pred. No. 2.8e-05;
Matches 185; Conservative 144; Mismatches 345; Indels 337; Gaps 42;

Qy 441 ILSHEISNGCKKALPQNSSLPTPIAK-----LEARGSKRQYHRAMQTEKHDNL 494
Dy 299 mlsehknfmimcnkn-ipqeelqtqwikewhgeflldhrsklpksckant----- 350
Qy 495 AQQIAQSQDVERHNSSTCVETLD-----AAKTKIQKVQVQENLHGMPPEVIEIE 543
Dy 351 -----lyeacekecidpmkyrdwiirskfewhtlskeyetqkvpkenaeay---likis 402
Qy 544 DDPTDGARK-----GKNTASISKAGSKGNSSPVKKTAKEKCIYPKTPAK 588
Dy 403 enkdakvsi1lnncdaesyksyckdhttlvksvlnqndntik---ekrehidlddfsk 459
Qy 589 KGRAGRKKSVPPPAHASEIOLWOPTTPKPTPLSRSKPKGKGRKSIQDSGKARGPSGELLQ 648
Dy 460 fg-cdksv-----tntkwecknpyi-----lstkd-----vc- 488
Qy 649 DSIAEIIYRMONLYLGDKEREQEONAMVLYKGDGALVPYESK---KRKPRPKVD-----ID 701
Dy 489 -----vprrqelcglndirlydknllmikehilaiaiesrilkrkyknkddkevckii 543
Qy 702 DET-----TRWN-----LLMGKDEKDEKDEKKEK-----WVEERR----- 737
Dy 544 nktfadiirdiigtgyndlnsrklygkintnskyvhrnkndklfrdewkvikkdawn 603
Qy 738 ---VPRGRADSFIAARMHLVQGDRRFSPWKSGSVVDVIGVFLTQNVSDHLSSSAFMSLAA 793
Dy 604 viswvfk---dktvckeddienipqffrfsewgddy----- 637
Qy 794 RFPKPLSSSREDERNVRSVVVE-----DPEGCILNLNEIPSWQEKVQHPDMEVSGVDS 847
```

```
Db 638 -----cqdktkmletikveckepecdndcksknsykewiskk----- 677
Qy 848 GSKEQLRDCNSGIERFNFLEKSIQNLEE-----EVLSSQDSFDPAIF-----QSCGRVGS 898
Db 678 -----eeynkqakqyqekgnnykmysefksikpevyllkkyseksninf 723
Qy 899 CSCSK-----SDAEFTTRC-ETKTV-----SGTSOSVOTGSPNLSDEICLOQNERPHLYE 948
Db 724 edefkeelhsdyknkctmcpvkdvpsilrnnedtsqeaapeenteia-hrtetpsise 782
Qy 949 G-SGDVQKQ-----ETTNAQKKPDLEKTMNWKDSVCFGQPR 984
Db 793 gpkgneqkerddslskisvspensrpetdaktstnllklkgvdism---pkavigssp 839
Qy 985 NDTNQTTPSSSYEQACATR-----QPHVLDIEDFGM-QGEGLYGYSWMSISPRV----- 1031
Db 840 ndninvteggdnisgvnksplsdvvpdkkeledqnsdeseetvvnhiskspsinnqdds 899
Qy 1032 -----DRVKNKNVPRFFRQGSVPRE 1053
Db 900 ggsatvsesssntglslddrngdtfvrtqtdtantedvirkenadkdedekgadeerh 959
Qy 1054 FTGOIIPSTPHELPGMLSGSS---SAVOEHQDDTQHNOQD-----EMNKASHLQKTF--- 1103
Db 960 stsesisspeekmlcdneggnslnhveevkehtsmdnvqggqgvinmnmvekeikdtlenp 1019
Qy 1104 ---LDLLSSESLTRQSTKQNIITDGLPRDRTAEDVVDPLSNSSSLQNILVESNSSNK 1160
Db 1020 sssldegkaheelsepnlsddqmsntpgpndtsetterisnneykvnredertitk 1079
Qy 1161 EQTAVEYKETNATILREM-KGTLAGKKPTSQWDSLRKDVE-----GNEGRO----- 1206
Db 1080 evedivlk---shmrresddgldydsldstvyndesedaekmkgndtsemshssghi 1136
Qy 1207 --ERNKNMNSD-----YEAIRASISEISEATKER 1236
Db 1137 esdqkndmktvgdltgthvneisvptgeideklreskeskhkaeeer 1187

RESULT 13
AAW22477
ID AAW22477 standard; Protein; 1435 AA.
XX AC AAW22477;
XX DT 11-SEP-1997 (first entry)
XX DE Silaic acid binding protein.
XX KW DBL gene family; SAPP; sialic acid binding protein; merozoite; malaria;
XX KW Duffy antigen binding protein; DABP; erythrocyte; var-1; var-2; var-3;
XX KW var-7; vaccine; therapy; immune response; Plasmodium.
XX OS Plasmodium falciparum.
XX PN W09640766-A2.
XX PD 19-DEC-1996.
XX PF 07-JUN-1996; 96WO-US09508.
XX PR 07-JUN-1995; 95US-0487826.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
XX PI Wellens TE;
XX WPI; 1997-052231/05.
XX DR N-PSDB; ANT72888.
XX PR New malaria vaccines - contains cysteine-rich DBL family protein
```

```
PT binding domains homologous domains of the Duffy and sialic acid
XX binding proteins
XX Example 1; Page 37-40; 96pp; English.
CC This sequence represents the full length sialic acid binding protein
CC (SAPP). SAPP and the Duffy antigen binding protein (DABP) are soluble
CC proteins that appear in the culture supernatant after infected
CC erythrocytes release merozoites. DABP and SAPP mediate the binding of
CC merozoites and schizonts to the erythrocyte surface. These proteins are
CC necessary for erythrocyte invasion by the parasite. This sequence can be
CC used in the compositions of the invention. The compositions are for the
CC treatment and prevention of malaria, and comprise either a nucleotide
CC sequence or encoded polypeptide of the var-1, var-2, var-3 or var-7 genes
CC of the DBL gene family, a family of genes having homology with conserved
CC regions of DABP and SAPP. The compositions are used for the treatment and
CC prevention of malaria. They are also used in the preparation of vaccines
CC for inducing a protective immune response in a mammal to Plasmodium
CC merozoites (especially Plasmodium falciparum or Plasmodium vivax).
XX Sequence 1435 AA;
SQ
Query Match 2.1%; Score 188.5; DB 18; Length 1435;
Best Local Similarity 18.3%; Pred. No. 2.8e-05;
Matches 185; Conservative 144; Mismatches 345; Indels 337; Gaps 42;
Qy 441 ILSHEISNGNGCKKALPQNSLTPIMAK-----LEEARGSKROYHRAMGOTEKHDLML 494
Db 299 mlsekhninnckn-ipqeelqitqwikewhgefillerdnrsklpkscknnt----- 350
Qy 495 AQOIAQSDVERHNSSTCVGYLD-----AAKTKIOKVQVENLHGMPPEVIEIE 543
Db 351 -----lyeacekecidpcmkrydwiirskfwhltiskeyetdkvpknaeny---likis 402
Qy 544 DDPDTGARK-----CKNTASISKASKGNSSPVKTAETKEKIVPKTPAK 588
Db 403 enkdakvslllnnccdaeyskycdckhttlvksvlnngndntik---ekrehldiddfsk 459
Qy 589 KGRAGRKKSVPPPAHASEIQLMQPTPPKTPLSRSKPKGKRSKIDSDSGKARGPSGELLCO 648
Db 460 fg--cdknsvd-----tntkvwecknpyl-----lstkd-----vc- 488
Qy 649 DSIAELIYRMQNLVLDKEREQEQNANWLYKGDGALVPYESK---KKKPRPKVD-----ID 701
Db 489 -----vpprrgeicignidrydknllmikelaiaiaiesrllkrkyknkdkkevckli 543
Qy 702 DET-----TRWN-----LLMGKGEKEGEDEKDKKEK-----WDEERR----- 737
Db 544 nktfadiirdiiggtaywndlsnrkivgkintaskyvhrrnkndklfrdewkvikkdwnv 603
Qy 738 -----VFRGRADSFIAARMHLVQGRDRFSPMKGSVVDPSVIGVFLTONVSDHLSSAFMSLAA 793
Db 604 viswvfk---dktvckeddienipqfrfwsewgdy----- 637
Qy 794 RPPPKLSSREDERNVRSVVE-----DPEGCIINLNEIPSWQEKVQHPDMEVSGVDS 847
Db 638 -----cqdktkmletikveckepecdndcksknsykewiskk----- 677
Qy 848 GSKEQLRDCNSGIERFNFLEKSIQNLEE-----EVLSSQDSFDPAIF-----QSCGRVGS 898
Db 678 -----eeynkqakqyqekgnnykmysefksikpevyllkkyseksninf 723
Qy 899 CSCSK-----SDAEFTTRC-ETKTV-----SGTSOSVOTGSPNLSDEICLOQNERPHLYE 948
Db 724 edefkeelhsdyknkctmcpvkdvpsilrnnedtsqeaapeenteia-hrtetpsise 782
Qy 949 G-SGDVQKQ-----ETTNAQKKPDLEKTMNWKDSVCFGQPR 984
Db 783 gpkgneqkerddslskisvspensrpetdaktstnllklkgvdism---pkavigssp 839
Qy 985 NDTNQTTPSSSYEQACATR-----QPHVLDIEDFGM-QGEGLYGYSWMSISPRV----- 1031
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:17 ; Search time 20.61 Seconds
(without alignments)
1689.960 Million cell updates/sec

Title: PCT-US01-13059-2
Perfect score: 9089
Sequence: 1 MQSIMSSAVNATEEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	210.5	2.3	1805	1 US-07-853-913-2	Sequence 2, Appli
2	195	2.1	1780	1 US-08-769-309A-5	Sequence 5, Appli
3	195	2.1	1780	3 US-08-994-570-5	Sequence 5, Appli
4	188.5	2.1	1435	2 US-08-568-459A-4	Sequence 4, Appli
5	188.5	2.1	1435	2 US-08-487-826B-4	Sequence 4, Appli
6	187.5	2.1	1346	2 US-08-635-121-2	Sequence 2, Appli
7	185.5	2.0	2409	6 5180808-2	Patent No. 5180808
8	185	2.0	2441	3 US-08-961-739-2	Sequence 2, Appli
9	183	2.0	2441	1 US-08-194-468-2	Sequence 2, Appli
10	182	2.0	1863	1 US-08-598-591-2	Sequence 2, Appli
11	182	2.0	1863	1 US-08-798-691-2	Sequence 2, Appli
12	182	2.0	1863	1 US-08-798-691-6	Sequence 6, Appli
13	182	2.0	1863	3 US-08-825-487A-2	Sequence 2, Appli
14	182	2.0	1863	3 US-08-825-487A-6	Sequence 6, Appli
15	182	2.0	1863	4 US-09-074-476-2	Sequence 2, Appli
16	182	2.0	1863	4 US-09-074-476-4	Sequence 4, Appli
17	179	2.0	1663	5 PCT-US93-07261-16	Sequence 16, Appli
18	177.5	2.0	3248	1 US-08-353-700-1	Sequence 1, Appli
19	177.5	2.0	3248	5 PCT-US95-16216-1	Sequence 1, Appli
20	177	1.9	1863	2 US-08-603-753D-2	Sequence 2, Appli
21	177	1.9	1863	4 US-09-099-753-2	Sequence 2, Appli
22	177	1.9	1863	4 US-08-986-106-2	Sequence 2, Appli
23	175	1.9	1210	3 US-08-320-559-26	Sequence 26, Appli
24	175	1.9	1210	3 US-08-545-860D-26	Sequence 26, Appli
25	175	1.9	1210	5 PCT-US94-04496-26	Sequence 26, Appli
26	174	1.9	1588	5 PCT-US93-07261-11	Sequence 11, Appli
27	173	1.9	2414	1 US-08-227-536-2	Sequence 2, Appli

28	173	1.9	2414	5 PCT-US95-04682-2	Sequence 2, Appli
29	172.5	1.9	1187	1 US-08-320-559-28	Sequence 28, Appli
30	172.5	1.9	1187	3 US-08-545-860D-28	Sequence 28, Appli
31	172.5	1.9	1187	5 PCT-US94-04496-28	Sequence 28, Appli
32	172.5	1.9	2842	1 US-07-741-940-7	Sequence 7, Appli
33	172.5	1.9	2842	1 US-08-289-548A-7	Sequence 7, Appli
34	172.5	1.9	2842	1 US-08-452-654-7	Sequence 7, Appli
35	172.5	1.9	2843	1 US-08-452-655B-2	Sequence 2, Appli
36	172.5	1.9	2843	1 US-08-452-655B-7	Sequence 7, Appli
37	172.5	1.9	2843	4 US-08-450-582-2	Sequence 2, Appli
38	172.5	1.9	2843	4 US-08-450-582-7	Sequence 7, Appli
39	172.5	1.9	2973	2 US-08-821-355A-7	Sequence 7, Appli
40	172.5	1.9	2973	2 US-09-003-687A-7	Sequence 7, Appli
41	172.5	1.9	2973	4 US-09-136-605-7	Sequence 7, Appli
42	172	1.9	1852	1 US-08-425-061-24	Sequence 24, Appli
43	172	1.9	1852	2 US-08-825-886-24	Sequence 24, Appli
44	172	1.9	1863	1 US-08-425-061-16	Sequence 16, Appli
45	172	1.9	1863	1 US-08-480-784-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-07-853-913-2
; Sequence 2, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853,913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1805 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-07-853-913-2

Query Match 2.3%; Score 210.5; DB 1; Length 1805;
Best Local Similarity 18.7%; Pred. No. 3.2e-08;
Matches 283; Conservative 216; Mismatches 564; Indels 447;

QY	60	PRKPAELPKVVVEGPKRKRKAATQEVKSK--ETSGAKKKNLKESATKKPANVGMSN	117
QY	118	KSPEVT--LKSCRKALNFDLENGPDARQDSESETVONSVCANSEFSEIRDAI--GGTNGSF	174
QY	211	RLSQAVRGAREC-----LEVOQLADRDLSQERREALERLEGRW	251
QY	175	LDSVSQIDKTN--GLGAMNOPLEVSMGNQPKLSTGAKLRADQOPLDLTRNQOCCFPVATQ	233
QY	252	QDRLOATDKFQLAVALAEQOGLOGSQIAQLLEGGOOLAH-----LKMSLSEVATY	303
QY	234	NTQFPMENOOAWLQMKNLQIFPGFNQOPRMTIRNOQCLAMGNOQPMYLGITPRPALVS	293
QY	304	RTLLEAENSR--LQTPGRGQASLGFLDPKL-----KPNF--LGIPEDQYL--GSVLPALSP	354
QY	294	GNQOLGGPOGNCRRP1--FLNHQTCPLPAGNOLYGSPTDMHQLVMSTGGQOHLKLNKQOQP	351
QY	355	TSFPSPLPNTLETPVTAFLKTOEFLQARTPTLAS--TPIPPISEAP-----CPPN	402
QY	352	SLIRGOOPCVPLIDQOATP---KGFTHLNOMVATSMSPG-----LR	391
QY	403	AEYRAQEVPLSLQTAPEPLWKATVPSSAILPELEPGKGQOQHFPDDLTSLATNLN	462
QY	392	PHSOSQVPTTYLHVESVSRILNNGTTCORSRAPAYDSLOQDIHQGNKYI-----LSHEI	446
QY	463	PHH-----PT-----LEAKOGESSESRVSI--FQDEQOQWELVEKEADIEVKV	505
QY	447	SNGNGCKKALPONSSLDPTIMAKLEEARGS---KROYHRAMGQTEKHDNLNLAQIAQSQ	502
QY	506	ENSSAQK---TQESGLDT-----EETQDSQPLQKETLKALG--EEPLMSLKIONYETA	554
QY	503	DYERUNSS1-----CVEYLDAAKTKIOKVVOENL-----HGMP--PEVIEIEDPTD	548
QY	555	GRKNCNSFEGHGLTGLEGKEKQIPLKSLEKNVSEKTLLENGVPVLSELLGKEDRTTE	614
QY	549	-----GARK-----GKNTASISKAGSKN-----SSPVKKT	574
QY	615	DQELMSPKGLKRFSSLCCKESQEVVYRPSKEGNLESWTAFKEESQHPGLFGPAEDQMLERL	674
QY	575	AEKEKCIVPKTPAKG---RAGRKKSVPPPAH-----ASEIQLMQPTPPK	616
QY	675	VEKEDQSPRSPEEEDQACRPLKQENOEPLGYEAEQOILERLIEKESLESRAPEED	734
QY	617	TPLSRSKPKGK---GRKSIODSGKARGPSGELLCCODSIAEIIYRMONLYLGDK--BREQE	671
QY	735	QZAGRSLOKQENOEPLGYEAEADOMLER-----LEKESQESLAKSPENORIGKPLRENQ	789
QY	672	QNAWVLYKGDGALVPYESSKKRKRPKVDIDDETTRINWNLGMKGDEKGD--EEKDKKKEK	730
QY	790	KSURLYEENQOETFVPLESRNORPLASLEVEEEOQIRVKPEKVSQDSLSGLSAEENQVPLR	849
QY	731	WHEEE---RRVFRGRADSFARHVLHVGDRRFPSPWKGVSDVSVIGVFL-----TONVS	780
QY	850	YLEEDDCTINKSLLEDKTHKSLGSLDRNGDSIIIP-----QESTQVSLRPPPEEEDQRIV	904
QY	781	DHLSSAFMSLAARPPPKLSSSREDERNV--RSYVVVEDPEGICILNLNEIPSWQEKVQHP	838
QY	905	NHLEKES-----QEFRSSEEEEOVMERSLEGENHE-----SLSSVEKEQDQWVE--S	949
QY	839	DNEVSGVDSG-----SKEOLRDCNSNGIERFNFLEKSLTONL-----EE	876
QY	950	QLEKESQDSGSKGSLSEDSQETGPLEKENAESLRSLAGODOEQKLEQTOOTLRAVNGEQ	1009
QY	877	EVLSSQDSFDP-----AIFQSCGRVSGCS--CSKSDAEFFPTR--CETKTVSGTS--	922

; INFORMATION FOR SEO ID NO: 5:

; SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids

```

; LENGTH: 1780 amino acids
; TYPE: amino acid

```

TYPE: AMINO ACID
TOPOLOGY: Linear

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-769-309A-5

Query Match	2.18:	Score 195:	DB 1:	Length 1780:
-------------	-------	------------	-------	--------------

Query Match 2.18; Score 193; DB 1;
Best Local Similarity 17.88; Pred. No. 6.5e-07;

Best Local Similarity	17.8%	Fied. NO: 6.3E-07;
Matches 332:	Conservative 262:	Mismatches 690:
		Indels 584:
		Gaps 84:

QV 105 ATKKPANVGDMNSKSP E V T L K S C R K A L N F D L E N P G D A R O G D S E S E I V O N S -----SGAN 158

QJ 103 81KRFANVGDMSNN3FEV1GRSCAAENFDEENFGDARQGDSSEIVQNS - - - - -SGAN 138

Db 23 AEPEPSGGGPSAEAPDTT-----ADPAIA-ASDPATKILOKNGOI,STINGVA 69

ZC ZW 01 000070869EDYI
ADFAIA ASDFAINDUQKNOQJINVO 03

QY 159 SFSEIRDAIGGTNGSFLDVSQIDKTNGLGAMNOPL-----VSMGNOPD 203

[illegible]

Db 70 EQDELSLQEGDLNG-----QKGALNGQGALNSQEEEEVIVTEVGQRDSEDEVSRDSDK 122

.....

QY 204 KLSTGAKLARDQQPDLTRNQQC-QFPVATQN-----TQFPMENQQAWLQMKN--QLIGFP 256

[illegible]

Db 123 EMATKSAVVDITDDGQEEENRNIQIPSSSENLEELTQ-PTESQANDIGFKKVFVGVFK 181

Qy 257 FGNNQPR-----MTIR-----NQPCLAMG-----NQQPMYLGTPRPA 290

[illegible]

Db 182 FTVKKDKTEKPD TVQLLTVKKDEGEAGAGD HQDPSLGAGEA ASKESEPKQ --STEKPE 239

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

QY 291 LVSGNQQLGGPQGKRRPIFLNHQTCLPAGNQLYGSPDMHQLVMSTGGQGHLLIKNQQP 350

240 ETI KPEO-----CHNPTD-----DABECOMVEEC-VBECCEVVEEEN 276

DB 240 ET'LKREQ-----SHAETSP-----PAESGQAVEEC---KEEGEERQEREP 2/6

351 GSIIRGQNPVPIIDQOPATPKGFTHI NOMVATSMSSPCI PRHSOSQVPTTYI HVPSVSP 410

QY 3JI GSLRGQQPCVFLIDQQFATPAGFTHLNQMVATSMSSPGLRPHSQSQVPTIYLHVESVSR 410

```
Db 277 SK--SAESPTSPVTSETGSTFKKF--FTOGWAGWRKKTSEPKPEDEFEAS-----323
```

DD 277 SK 3AESF1SFVISEIGSIFRNF - FIQWAGWKKRKRISFRKFRNEDEVEAS - - - - - 323

QY 411 ILNGTGTQSRAPAYDSLOODIHOGNKYILSHEISNGNGCKKALPONSSSLPTPIMAKL 470

[illegible]

Db 324 -----EKKKEQEPEKVDTEED--GKAEVASEKLTASEQAHPQEPAESAHPEPRLSAEY 373

Qy 471 EEAR-GSKRQYHRAMGQTEKHDLNLAQQIAQSQDVERHNSS-----TCVEYLDAAKKT 522

—
..

—
..

—
..
..
..
—

..
..
—
..
—

..
..
—
..
—

..
—

Db 374 EKVELPSEEQVSGSGPSEKPAPlATEVF-DEKIEVHQEEVVAEVHSTVEERTEEQKT 432

RESULT 3

RESUL³
US-08-994-570-5

US-08-994-370-3
: sequence 5. Application us/08994570; sequence 3, Applicant
: Patent No. 6090929

GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Nauert, Brian J.,
APPLICANT: Klauck, Theresa M.
TITLE OF INVENTION: Protein Binding Domains of Gravin
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/994,570
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: No. 60929and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-994-570-5

Query Match 2.1% Score 195; DB 3; Length 1780;

Best Local Similarity 17.8%; Pred. No. 6.5e-07;
Matches 332; Conservative 262; Mismatches 690; Indels 584; Gaps 84;

QY 105 ATKPANVDMNSKSPVLTLSKRCALNFDLENPCDARGDSESEIVQNS-----SCAN 158
DB 23 AEPEPGSGSAEAPDT-----ADPAIA-ASDPATKLLQKNGQLSTINGVA 69
QY 159 SFSEIRDAIGTNGSFLDSVSQIDKTNGLGAMNOPLE-----VSMGNQPD 203
DB 70 BDELSLQEGDLG-----QKALNGQAGALNSQEEVIVTEVGQDSDVSERDSK 122
QY 204 KLSTGAKLARDQPDLLTRNOOC-QFPVATON-----TOFPMENOQAWLOMKN--OLIGFP 256
DB 123 EMATKSAVVDITDDGQENRNIQIFSSSNLEELTQ-PTESQANDIGKFKVFGFK 181
QY 257 FGNOQPR-----MTIR-----NQOPCLAMG-----NQOPMYLIGTRPA 290
DB 182 FTVKDKTEKPDVTOLLTVKKDEGEAGAAGDHODPSILGAGEAASKESEPAQ--STEKPE 239
QY 291 LVSGNQQLGGQGNKRFILNHTCLPAGNOLYSPDMDHOLYMTSGQOQHLLKNOQP 350
DB 240 ETLKREQ-----SHAISIP-----PAESGOAVEC--KEEGEEKQKEP 276
QY 351 GSLIRGOQPCVPLIDQOPATPKGFTHLNQMVATSMSSPGLRPHSPQSQPTTYLHVESVSR 410
DB 277 SK--SAESPTSPVTEGTFKKF--FTQWAGWRKKTSTFKPKDEVEAS-----323
QY 411 ILNGTTGTQCRAPAYDSLOQDTHQGNKYLSILSINGNGCKKALPONSLSPTPIMAKL 470
DB 324 -----EKKKEQEPKVDTEED---GKAEVASEKLTASEQAHPQEAESAHEPRLSAEY 373
QY 471 EEAR-GSKROYHRAWGOTERHDLNLAQOIAOSODVERHNS-----TCVEYLDAAKT 522
DB 11

DB 374 EKVELPSEEQVSGSQPSEKPAPLATEVF-DEKIEVHQEEVAVAEHVSTVEERTEQOKT 432
QY 523 KIQK-----VVOENLHGMPEVIEIE-----DDPTDGARKGKNTASISKGA 563
DB 433 EVEETAGSVPAEELVGMDAEPQEAEPKELVKLKETCVSGEDPTQGDLSPEKVLSP-P 491
QY 564 SKGNSSPVKTAETAEKCIIVPKTPAK-----KGRAGRKKSVPPEAHASIQWLQWPTTP 615
DB 492 PEGVSEVEMLSQERMKVQGSPLKLTSTGLAKLKKQKQKRGGGDESGHTQVPA 551
QY 616 KTLPSRKPKGRKKSQDSGKARGPSGELLIC-QDSIAEIIYRMONLYLGDKEREQONA 674
DB 552 DSPDSQEQKQESSASSPEEPE-----EITCLEKGLAEV-----QODGEAEEGA 595
QY 675 MVLKGDG-----ALVPYES-----KKRPRP-KVIDIDETTRINLLMG-----713
DB 596 ---TSDGEKKREGVTPWASFKNMVTPKKRVRRPSESDKEDELKVKRSATLSTESTASE 651
QY 714 ---KGDEKGEDEKDKK---EKWE-----EER 736
DB 652 MQEEMKGSVEEPKPEEPKRVDTSVSWEALICVSGSSKKRRRRSSDEEGPKAMGGDHQ 711
QY 737 RVFRGRADSFIAHMLYQDRRFSPKWSVVDVSVIGVFLTONVSDHLSS-SAFMSLAARF 795
DB 712 KADEAGKDKETGTDILAGSOEHDPGOGSSSPQAG---SPTGEGYSTWESFKRLVT--766
QY 796 PPKLSSSREDERNVRSVV-----813
DB 767 PRKKSKELEKSEDSIAGSGVHSTPDTEPGKEESVSIKKFTIPGRKKRPDKQKQOAP 826
QY 814 VED---PEGCIIILNLEIPSWQEKVQHPSDMEVSGVDSKQELR-DCSNSGIERENFLEKS 870
DB 827 VEDAGPTGA---NEDDS---DVPVAVPLSEYDAVEREKMEAAQQAQKGAQEP--QKA 875
QY 871 IONLEEEVLSQDSFDPALFOSCGRVGSCSKSDAEFFPTTRCTETKTVSGTSQSVQSGSP 930
DB 876 ATEVSKELSESQ-----VHMAAAVADGTAAATIIERSPSWISASVTEPLE 922
QY 931 NLSDEICLOGNE-----RPHLYEGSD--VOKQETTNAQKKPD---967
DB 923 QVEAEAAALLTEEVLEREVIABEEPTVTPELPENREARGDTTVVSEALTEAVTAATAG 982
QY 968 -----LEKTNMKDVCFCGPRNDTNMQTTTPSSSYEQCATRPHVLDIEDFGM 1015
DB 983 PLGSEECTEASAAEETEMVSQVLSQDTPDTEATPVQVE-----GGVPDIEQER 1036
QY 1016 QGEGGLYSWMSISPRVDRVKNKNVPRFRFGGSGVPREFTGQIIPSTPHELPGMGLSGSS 1075
DB 1037 RTQEV-----LOAVAEEKVEES-----QLPGTG--GPE 1062
QY 1076 SAVQEHODDTQHNQODENMKASHLOK-----1101
DB 1063 DVLOPVO-RAEAERPEQAEASGLKKTEDVVLKVDQAEKTEPTQGVKVGQTTPESEK 1121
QY 1102 --TFDLNLSSECLTRQSTT-----KQNTDGLPRDRTAEDVDVPLSNSSQLNLY 1153
DB 1122 APQVTESSIESSELVTTQAEETLAGVKSOEMVMEQAIIPD-SVETPTDSETDGTSPVADF 1180
QY 1154 ESNSSNKEQTAVEKKTENATILREMKTADGKKPTSDWLSRKVDVSGNEGRQENK--N 1211
DB 1181 APGTTQDE-IVEIHEENEVHLVPRGTEAEAVPAQKERPAPSSFFVQETKEQSKMED 1239
QY 1212 NMSIDYE-AIRRAISIEISEAIKERMNNNLAVRIKD--FLERIVKDHGID--LEWLR 1266
DB 1240 TLEHTDKESVEITVSILSKTEGTQE--ADQVADEKTKDVPFEGL---EGSIDGTIVSR 1294
QY 1267 ESPDPKADKYLILSIRGLGKSEVCR--LITLHNLAPPVDTNVGR-IAYRWGVPLOPLP 1323
DB 1295 EKYTE-----VALKGEGETEAECKDDALELQSHAKSPSPVEREMVVOVEREKTEAP 1348
QY 1324 ESLQLHLLLE--LYPVLESIQKFLWPRCLKDQRTLVHLVOLITFGK-----1368
DB 1349 THVNEEKLHETAVTVSEEVSKQI-----LOTVNVPIIDGAKEVSSLEGSPPPC 1397

Db 1137 ESDQKNDMKTVDLGTTHVQNEISVPVTGEIDKLRKESKSIHKAEEER 1187

RESULT 5

US-08-487-826B-4

; Sequence 4, Application US/08487826B

; Patent No. 5993827

; GENERAL INFORMATION:

; APPLICANT: Sim, Kim L.

; APPLICANT: Chitnis, Chetan

; APPLICANT: Miller, Louis H.

; APPLICANT: Peterson, David S.

; APPLICANT: Su, Xin-zhaun

; APPLICANT: Wellem, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobe Martens Olson & Bear

; STREET: 620 Newport Center Drive 16th Floor

; CITY: Newport Beach

; STATE: California

; COUNTRY: US

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/487,826B

; FILING DATE: 10-SEP-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Israelsen, Ned

; REGISTRATION NUMBER: 29,655

; REFERENCE/DOCKET NUMBER: NIH121.001CP1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1435 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHEICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: Plasmodium falciparum

; US-08-487-826B-4

Query Match

Best Local Similarity 2.1%; Score 188.5; DB 2; Length 1435;

Matches 185; Conservative 144; Mismatches 345; Indels 337; Gaps 42;

Qy 441 ILSHEISNGCKKALPQNSLPPIMAK-----LEEARGSKROYHRAMGQTEKHDLNL 494

Db 299 MLSEKNNINCKN-IQEEELQITQWIKWEHGEFLLEDNRNSKLPKSKCKNNT----- 350

Qy 495 AQQTAQSDVHRHNSSTCVYLD-----AAKTKIQKVQVQNLHGMPPVEIEIE 543

Db 351 -----LYEACEKECIDCKMYRDIIRSKFEWHTLSKEYETQKVPKNAENY---LIKIS 402

Qy 544 DDPTDGARK-----GKTAISKASKGNSPVKKTAEKCIIVPKTPAK 588

Db 403 ENKNDKAVSLLNNCDAEYSKYCDCKHTTTLVSVLNGNDNTIK---EKREHIDLDDFSK 459

Qy 589 KGRAGRKSVPPPAHASEIOLWQPTPPKTPLSRSPKPKGKRSIQDSKGKARGPSGELLQ 648

Db 460 FG--CDKNSVD-----TNTKWECKNPII-----LSTKD-----VC- 488

Qy 649 DSAETIIRMONLYLGDKEREQEONAMVLYKGDGALVPYESK---KRKPRPKVD-----ID 701

Db 489 -----VPPRQELCLGNIDRIYDKNLLMIKEHILAIAYESRIILKRYKNKDKVECKII 543

Qy 702 DET-----TRIWN-----LLMGKDEKEDEKDKKKEK-----WWEERR----- 737

Db 544 NKTADIRDIIGGTDYWNLDLSNRKLVGKINTNSKYVHRNKNKDLFRDEWVKVKKDVWN 603

Qy 738 ---VERGRADSEIARMHLVGGDRFSPWKGSVVDSVIGVFLTONVSDHLLSSSAFMSLAA 793

Db 604 VISWVFK---DKTVCKEDDIENIPQFFRFSEWGGDY----- 637

Qy 794 RFPKLSRREDERNVRVVVE-----DPEGCILNLEIPSWOEKVQHPFSDMEVSGVDS 847

Db 638 -----CQDKTKMIETLAVECKEPCEDDNCCKSKSYKWKISKK----- 677

Qy 848 GSKQLRDCSNSGIERNFLEKSTONLEE-----EVLSSQDSFDPALF-----QSCGRVGS 898

Db 678 -----EYNKQAKQYQYQKGNVKNMYSEFKSIKPEVYLKRYSEKCSNLNF 723

Qy 899 CSCSK---SDAEPPTTRC-ETKTV-----SGTSOSVOTGSPNLSDEICLQGNRPHLYE 948

Db 724 EDEFKEELHSDYKNKCTMCPEVKDVPISIIRNNEQTQSEAVPEPTEIHA-HRTETPISSE 782

Qy 949 G-SGDVQKQ-----ETNTVAQKKPDLEKTMNMKDSYCFQGP 984

Db 783 GPKGNEQKERDDDSLSKISVSPENSREPETDAKDTSNLLKLGVDVISM---PKAVIGSSP 839

Qy 985 NDNWQTPSSSYEQCATR-----QPHVLDIEDFGM-QGEGLYSWMSSISPRV----- 1031

Db 840 NDNINVTQGDNITSGVNSKPLSDDDRPKKELEDQNSDESEETVYVNHISKSPSINNGDDS 899

Qy 1032 -----DRVKNKNVPRFRFGGSGVPRE 1053

Db 900 GSGSATVSESSSNTGLSIDDRNGDTFVRTQDTANTEDVIRKENADKDEKADGERH 959

Qy 1054 FTQIIPSTPHELPGMGLSGSS---SAVQEHQDDTQHNOQD---EMNKASHLQKTF--- 1103

Db 960 STSELSPEEKMLTDNEGNSLNHEBVKHTSNSDNVQSGGIVNMNVEKELKDTLENP 1019

Qy 1104 --LDLLNSSEECITROSSTKQNTDCCLPRDRTAEDVDPPLSNSSLONLIVESNSNK 1160

Db 1020 SSSLDEGKAHEELSEPNLSDDQMSNTPGPLDNTSEETTERISNNEYKYNREDERTLTK 1079

Qy 1161 EQTAVEYKETNATILREM-KGTLDAGKKPTSQWDSLKQDVE---GNEGRO----- 1206

Db 1080 EYEDIVLK---SHMNRSDDGELYDENSIDLSTVNDSEDEDAFAKMGKNDTSEMHSNHOI 1136

Qy 1207 --ERNKNMDSID-----YEAIRRASISEISEAIKER 1236

Db 1137 ESDQKNDMKTVDLGTTHVQNEISVPVTGEIDKLRKESKSIHKAEEER 1187

RESULT 6

US-08-635-121-2

; Sequence 2, Application US/08635121

; Patent No. 5910442

; GENERAL INFORMATION:

; APPLICANT: Gelman, Irwin H.

; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10112-0228

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 1.5

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,121
FILING DATE: 19-APRIL-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Richard S
REGISTRATION NUMBER: 26,154
REFERENCE/DOCKET NUMBER: A30558 - 165/33603
TELEPHONE: 212-408-2558
TELEFAX: 212-765-2519
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1346 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-635-121-2

Query Match 2.1%; Score 187.5; DB 2; Length 1346;
Best Local Similarity 17.5%; Pred. No. 1.7e-06;
Matches 278; Conservative 192; Mismatches 513; Indels 607; Gaps 70;

QY 353 LIRGQPCV-----PLIDQAPATPKGFTHLNMVATS-----MSS 387
DB 63 LKMSREMCVSGDHTQLDLSPEETLPHKPGIYSEVEMLSQERIKVQGSPLAKLFSS 122
QY 388 PGLRPHS-----QSQPTTYLHVESVSRILNGTTGTCQSRAPAYDSLOODTHQ 436
DB 123 SGLKLSGKKQKRGKGDEPFGYQIHITES-----PESADEQK 163
QY 437 GNKYILSHEISNGCKKALPONSSILPPIAKL--EEARGSKROYHRAMGQTEKHDLNL 494
DB 164 GESSASSPEETTCLEKGP-----LEAPMGKRLKELLRGKKR-----KDHSLGI 211
QY 495 AQIAQSDVERHNSSTCVLEYLDAAKTKIKQVQENLHGMPEVIEIDD-----545
DB 212 LQKDGDTQETVRRPS-----ESDKEELEKVKSATLSSTDSYSEMDQEVKTVGEQK 264
QY 546 PTDGARKKNTAS-----TSKGASK-----GNSSPVKKTAKEKCK---580
DB 265 PEPKRRYDTSVSWPALICVGSKKRARKASSSDIRGPTLGGGQSGRGGOQRORSRTD 324
QY 581 IVPKTPAKKGRAGKKSVPV---PAHASEIQLWQP-----TPPKTPLSRSPKGGKRSKI 632
DB 325 AVPASTQEQDQAGSSPEPAGSPGEGVSWTEFKRLVTPRKKSLEKEAGRTLV 384
QY 633 QDSG-----KARGSGGELLCCODSTAEIIRYQNLVLDGK-EREQEQNAM 675
DB 385 VGAGCPLRSNRVEKNLGPFLRNSPDG-----GRKGGMGRQEQAT 424
QY 676 VLYKGDGALVPYESKKRPRKVDIDETTRIWNLLMGKGEKEDEKDKKKKEKWBEE 735
DB 425 VEDSG-----PVEINEDEP-----DVPVAVPL-----SEYDAVEREKME-----458
QY 736 RVFRGRADSFIAHMLVGGDRRFPKSGSVVDVIGVFLQNVSDHL--SSSAPMSLAA 793
DB 459 -----AQGNALPSCVGVVSEELSKTLVHTVSVAVIDGTRAVTSVEE 501
QY 794 RFPPLKSSS-----RDENRNRSVVVED-----PEGCIILNLNIPSWQ 831
DB 502 RSPSWISASVTEPLEHTAGEAMPVVEETKDIIAEETPVLQTQLPEG-----549

QY 832 EKQVHPSDMEVSGVD-----SGSKE-----OLRDCSNS 859
DB 550 -KDAH-DDMVTSEVDFTSEVATATSETSEALRTEEVTEASGAEETDMVSVAVSLDSDPT 607
QY 860 GIER--FNFLEKSIQNLLEEVLSSQDSFPALFQSGRGVSGSCSKSDAEFFTRCETKT 917
DB 608 TEETAPVQVEGGVLDTEEEERQTQ-----ALQAV-----ADKYKEESOVPATQTVQRT 657
QY 918 VSGTSQSVQOTGSPNLSDEICLOGNERPHLYEGSDVQKQOETTNVAQKPDLEKTMNWKDS 977
DB 658 GSKALEKVE-----EVEEDSEVLASEKEKDMPKGPVQEAEGAHLAQ-----699
QY 978 VCFGQPRNDTNQWTPSS-----SYEQCATRQPHVLDIEDFGMOGBGLGYSWMSISP 1029
DB 700 -----GSETQATPESLEVPETADVHVATCQ--VIKLOQLMEQ-----AVAP 741
QY 1030 RVDVRKNKNVPRFRFGGSGVPRFTGTQIIPSTPHELPGMGLSGSSSAVQEHQDDTHQ 1089
DB 742 ESSETLIDS-----ETNGSTP-----LADSDTADGTQDDTIDSQ 776
QY 1090 QDEMKNASHLOKTFDLNLSSEECITROSSTKONTIDGCLPRDRTAE-----DVVDPLSN 1145
DB 777 DSKATAAVRQSOV-----TEEAATAQKEPSTLPNNVPAQEEHGEPEGRDVLPTQOE 830
QY 1146 SSLQNLVYESNNSKEQTAVYEKTNATILREMKGTLDGKK-----PTSQ--1191
DB 831 LAAAAVPVWQKTEVGQEGEVDW-----LDGEKVKEEQEVFVHSGPNSOKA 875
QY 1192 ----WDLRKDVEGNEGROERNKNMNDSDIDEA-----IRRASISEIAKEKGMNM 1241
DB 876 ADVTIDS-----EYGVAGQCKEKESTEVSLSLEEGEMETDVEKEKRETRPEQVSEGEQT 932
QY 1242 LAVRIKDFLERIVKHGGIDLEWLRESPPDKAKDYLLSIRGLGLKSVCEVRLTLHLNL--1299
DB 933 AA-----PEHERNYGKPVLTLDMPSE--RCKALGSLG 964
QY 1300 --AFPVDNVGRIAVRMGWVPLQPLPESLQLHLLLELYPVLESIQKFLPRLCKLDQRTLY 1357
DB 965 SPSLPDQDKAGCIEQV-----QSLDTTVTQ---TAAVEKVI-----999
QY 1358 ELHYQLITFGKVFCKSPNCNACPMRCECRHFASAYASARLALPAPERLSLTA-----1412
DB 1000 ----ETVWISE---TGESPEC-----VGAHL-LPA-EKSSATGGHWTILQ 1034
QY 1413 ----TIPVPPESPFPVAPMIELPLLEKSL-----ASCAPSNRENCEPIIEEPASP---1460
DB 1035 HAEDTVPLGPES-QAESIPILVTPAP-ESTLHPDLQGEISASORERSEREDKPDAGPAD 1092
QY 1461 GOECTEI-----TESDIEDAYNEDPDEIPTIKLNIQFGMT-----1497
DB 1093 GKESTAIKVLKAEPEILELESKNKIVLNVIQTAVDQFARTETAPETHAYDSOTQVPAM 1152
QY 1498 -----LRHEMERNMELQEDMSKALVALHPTTISI-----1527
DB 1153 RLDSREPNCRTKMKVAKMHPVPOPREDQLVLTVLEAWLSSEMLAALAVESAGVKVSTIE 1212
QY 1528 ---PTPKLKNISRLRTEH-----QVYELPDSHRLDGMKREPDPSPVYLLAINTPGETA 1579
DB 1213 KLPPQPKDQK-----EHAADGQPLQSLAQAEAVSGNLTKESPDTNGPKL-----1256
QY 1580 NSAQPEQKCGGK--ASGKMPDETCE-----CNLSRE-----ANSTVTRGT 1620
DB 1257 -----TEERCPOKLSRKKKCLPSQSKRTRPQKRKTRCSQRETNQWPKMLVHACTSVRP-1310
QY 1621 LLIPCRAMRGSFPLNGTYFQVNLFPADHE 1650
DB 1311 ---ECENKSNKMLLGLPWTKISEPMRSRE 1337

RESULT 7
5180808-2
; Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.

;; CURRENT FILING DATE: 1997-10-31
;; EARLIER APPLICATION NUMBER: US 194,468
;; EARLIER FILING DATE: 1994-02-10
;; NUMBER OF SEQ ID NOS: 4
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 2
;; LENGTH: 2441
;; TYPE: PRT
;; ORGANISM: Mus
;; FEATURE:
;; NAME/KEY: VARIANT
;; LOCATION: (1)-(2441)
;; OTHER INFORMATION: xaa = Any Amino Acid
US-08-961-739-2

Query Match 2.0%; Score 185; DB 3; Length 2441;

Best Local Similarity 17.5%; Pred No. 7.9e-06;

Matches 342; Conservative 222; Mismatches 685; Indels 708; Gaps 84;

QY 5 MDSSAVNATEATEQNDGSRQDVLFDLNT-----PQKPKSKRKRKFMKPVVVEGPKRK 59
DB 287 MGATGVNPOLASKQSMVNSLPAPPTDINKNTSVTPVNMSQLQTSVGIVPTQAIATGPTAD 346
QY 60 P--RPAELPKVVVEGPKRKPKKAATQEKVSKETGSAKKNLKESAT-----KKPANVG 113
DB 347 PEKRLIQOQLVLLHLAHKCORREANGEXACSLPHCRMTNVLNLMHMHCOAPKACQVA 406
QY 114 DMSNKSPEVT-LKSCRK---ALNFDLENPGDARQGD---SESEIVONSSGANSFSEIRD 165
DB 407 HCASSRQIISHKNTCRHDCVCLPLKNASDRNQTILGSPASQIONTIGS-----V 459
QY 166 AIGGTNGSFLDSVSQIDKTN-----GLGAMNQPLEY----- 196
DB 460 GAGQONATSLSNPNIDPSSMRAYAAALGLPMYNQPTQLQPVGQQPAQPPAHQOMRT 519
QY 197 --SMGNQPKLSTGAKLARDQPDLLTRNQOQFPVATONTOPFMBNOQAWLQKNOLLIG 254
DB 520 LNALGNMSPVAGG-ITTDQOQPNLISESALPTSIGATN---PLMND-----G 564
QY 255 PFGNQOQPMTRNQOQCLAMGNQOPMY-----LIGTPRAL----- 291
DB 565 SNSNGIGSLSTIPTAAPSPSTGVKRGWHEHVTDLRSLHLVHLVQIAIFTPDPAALKDRR 624
QY 292 -----MENLAYAKKVEGDMYESANSRDEYHLLAEKIYKIQKELEKRRTRLKHQILGNOPAL 684
QY 300 GPQGNKRPIFLNHTCLPAGNOLYGSPTDMHOLVMSSTGGQOHLIKKNOQPGSLIRGQOP 359
DB 685 PASGAQPPVIPPAAQSVRRPNEGEL---PLPVNRMOVSGQ-----MNSFNPMSLGNVQLP 734
QY 360 CVPLIDQOQPATPKGTHLNQMV-ATSMSS-PG-LRPHSQSQVPTIYLHVESVSRILNGT 415
DB 735 -----QAPMGFRAASPMHNSVQMSMASVPGMATSPSRMPQPPN-----MMGT 777
QY 416 TGTQSRAPAYDSIQODITHQNKYILSHEISNGCKKALPONSLSLPTPIMAKLEEARG 475
DB 778 HANNIMAAQPTQNF-----LPQN----- 796
QY 476 SKROYHRAMQTEKHDLNLAQOIAQSDQYVERHNSSTCVE-----YLDAAKTKIKQVQE 530
DB 797 ---QPFSSGAMSVNSVGMGQPAAGVSGQGEPAALPNLMLAPOLPCPPVPTOS 853
QY 531 NLHGMPPEVIEEDDP-----TDGARKGKNTASISKASK 565
DB 854 PLHPTPPASTAAGWPSLQHTAPGMTPPQAPATQPTPVSQGTPTPTPGSVFSAAGT 913
QY 566 GNSSPYKTAKEKICIVPKTAPKAGRGRKKSVPVPAHASEIQLMQPT-----PKPTPLS 620
DB 914 QSTPTVQAAAQAQVTPQPTPVQ-----PPSVATPQSSQQOQPTPVHTQPGCTPLS 963
QY 621 RSK-----PKGGRKSIQDSGKARGPSGELLQCDQSIAEIIYRMQNLVYLGDKEREQON 673

DB 964 QAAASIDNRVPTPTSTVTSAAETSSQPGDPVPM-----EMKEVQT-----DAAPE----- 1010
QY 674 AMVLYKGDGALVPYESKRRKPRKPKVDIDDETTRINWLLMGKDEKDEKDKKKEKWE 733
DB 1011 -----PTES-KGEPRSEMEED-----LAGSSQVKEETDTTEQKSPMEV 1049
QY 734 EERRVFRGRADSFIAHMLVQDGRFSPWKSVVDSVIGVFLTQNVSDHLSSAFMSLAA 793
DB 1050 EEK-----PEVKEAKEEENSNDTASOSTSP 1078
QY 794 RPPPKLSSSREDER-----NVRSVVVEDPEGCIL-----NLNEIPSWQEKVQHPSDMEV 842
DB 1079 SQPRKKIKFPEELROALMPTLEALYRQDPESLPFRQVDPQLLGPDYDIYKVNPMDLST 1138
QY 843 --SGVDSGSKEQ-----LRDCSNSGIERFNFLEKSIONLEEEVLSQDS 884
DB 1139 IFRKLDTGQYQBPWQYVDDVRLMFNNALYNRKTSRV--YKCSKLAEFQEIEI----- 1190
QY 885 FDPALFQSCGRVYSGSCSKSDAEFPTRTCETKTVSTQSVOTGSPNLSDIEIC-----L 938
DB 1191 -DP-VMQS---LGYCCGRKYEFSPOTLCCYKQLCTIPRDAAYVSYQNYHFCGKCFTEI 1245
QY 939 QGNERPHLYEGSDGVQKQETTNNVQAKKPDLEKTMN----- 973
DB 1246 QG-ENVTL---GDDPSQPQTIS---KQDFEKKKNDTLDPEFVDCCKEGRKMHQICVLH 1298
QY 974 ---WKDS-VC-----FGQPRNDTNW---QTPPSSSY---EQCATRQPHVLIDIED 1012
DB 1299 YDIIWPSGFVCDNCKKTGPRKENKFSAKRLQTTFLGNHLEDVYNKFLRRQNHPEAGEV 1358
QY 1013 F-----GMOGEGLYSWSISPRVDRVKNKNVPRR-----FFRQGSVPR 1052
DB 1359 FVRVWASSDKTVEVKPMKSRFVDSGEMS-----ESFPYRTKALFAFEIDGVV 1408
QY 1053 ERTGOIIPST---PHELPGMG-----LSGSSSAVOR----- 1080
DB 1409 CFFGMHVQDTALIAPHQIOGCVYISYLDIHFPRPRCLRTAVYHEILIGYLEVKKLVV 1468
QY 1081 -----HODD---TOHQOODEMNKASHLQ-----KTFDLNLSSEECTRQSS 1119
DB 1469 TAHIWACPPSEGDDYIFHCHPPDQKIPKRLQEWYKMLDKAFARIINDYKDFKQAN 1528
QY 1120 TKONITDQCLPDRDRAEDVVDPLSNNSQNLILVESNNSKQETAVEYKETNATILREMK 1179
DB 1529 -----EDRLTSAKELPYFEGDFWPNVLEESIKELEQEEERKEEESTAASETPE 1577
QY 1180 GTLADGKKPTSQWDSLRKDOVEGNEGRQERNKNMDSIDYEAIRRASISSEISAIRK--- 1236
DB 1578 GSQGSK-----NAKKNNKTKNKK-----SSISRANKKKPSM 1611
QY 1237 -GMNNMLAVRIKDFLER-----IVKDHGGIDLEWLRESPPDKADYLSIRGLGLKSVE 1289
DB 1612 PNVSDNLSQKLYATMEKHKEVEFVHLAG---PVISTQPPVDPDPLLS----- 1658
QY 1290 CVRLTLHLNLAAPPVDTNVGR-----IARVWGVPLOPLPESLQHLLELYPVLESTQKFL 1344
DB 1659 -----CDLMDGRDAFLTLARDKH-----EFSSLRRSK 1686
QY 1345 WPRLCKLDORTLYELHYQLITFGK---VFCTKSRPNACPMRGECRHFASAVASARLAL 1401
DB 1687 WSTLC---MLVELHTQ---GQDRFVY-----TCNECKKHVETRWHTCTVEDYDLCI 1731
QY 1402 PAPEERSLTSATIPVPPEFPPVAIPMIELPL-L-EKSLASGAPSNRENCEPIIEEPASP 1460
DB 1732 NCYNTKSHTH-----KWKWGLGLDDGSSQGEQPSK-----SP 1765
QY 1461 GQECFEIT-----ESDIEDAYYNEDPDEPTIKLNEIQFGMTLREHMERMELOEQ----- 1511
DB 1766 -QESRRLSIQRQISLVHACQCRNANCSLPSC-----QMKRVVQHTKCKRK 1812
QY 1512 -----DMSKALVAL-----HPTTTSIPPP---KLK-NISRLRTEHOVYELPDSHRLDQ 1556

```

Db 1813 TNGGCPVCKOLIALCCYHAKHCENKCPVPFCLNKHNVQQOIOHCLQOAOIMRRMAT 1872
QY 1557 MDRK---EPDDPSYLLAIWTGCE---TANSAPQPEQ 1587
   | : | : | | | | | | | : | | | |
Db 1873 MNTNRVQOQLSPSPTSNAPCTPTQOQSTPTQTQPPAQ 1909
   | : | : | | | | | | | : | | | |

RESULT 9
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; TITLE OF INVENTION: RESPONSIVE GENES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9672
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)-546-4737
; TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-194-468-2

Query Match 2.0%; Score 183; DB 1; Length 2441;
Best Local Similarity 17.5%; Pred No. 1,2e-05;
Matches 342; Conservative 222; Mismatches 685; Indels 708; Gaps 84;

QY 5 MDSSAYNATEATEQNDGSRQDVFLEFLNKT-----PQKPKSRKRKFMKPVVVEGKPKRK 59
   | : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 287 MGATGVNPQLASKQSMVNSLPAPFTDIKNTSVTPNMSQLQTSVGIVPTQAIATGPTAD 346
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 60 P--RKAETLPKVVEGKPKRKPKKAATQKVKSKETGSAKKANLESAT----KKPANVG 113
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 347 PEKRKLIQQOVLVLLHAHKCRREQANGVRACSLPHCRTMKNVLNHNHTHCOAPKACQVA 406
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 114 DMSNKSPEVT-LKSCRK---ALNFDLENPGDARQGD---SESELVQNSSGANSFSEIRD 165
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 407 HCASSRQILSHWNKCTRHDPCVCLPLKNASDRNQOTILGSPASQIQTIGS-----V 459
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 166 AIGGTNGSFSDVSQIDKTN-----GLGAMNQPLEV----- 196
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 460 GAGQQQATSLSPNPIDPSSMRAYAAALGLPYMNQPTQLOQPOVFPQCPQAPPAHQMQRT 519
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 197 --SMGNQPKLSTGAKLARDQQPDLLTRNQOQCFVATNTQTFPMENQAWLQKNQLIG 254
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 520 LNALGNMNSVPVAGG-ITTDQQPPNLISESALPTSLGATN---PLMND-----G 564
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Db 1469 TAIWACPPSEDDYIFCHCHPPDQIPKPKRQEWYKMLDRAFAERIINDYKDFKQAN 1528
QY 1120 TKQNTIDGCLPRDRAEDVVDPLNSNSSQLNIVESNSSNKQTAVEYKETTATILREMK 1179
Db 1529 -----EDRLTSAKELPYFEGDFWPNVLESIRELEQEEERKKEESTAASETP 1577
QY 1180 GYLADCKPPTSOWDSLKRDVEGNEGRQENKNNMSIDYEATRASISISEAIKER --- 1236
Db 1578 GSQGSK -----NAKKNNKTKNKK-----SSISRANKKTPSM 1611
QY 1237 -GMNNMLAVRIKDELER-----IVKDHGGIDLEWLRESPPDKAKDYLLSIRGLKSV 1289
Db 1612 PNVSNLDSOKLATWEKKHEVFVILHAG---PVISTOPPIVDPDPLLS----- 1658
QY 1290 CVRLTLHLNLAAPPVDTNVR-----IAYRMGWVPLQPLPESLQLHLELYPVLESTQKFL 1344
Db 1659 -----CDLMGRDAFLTLARDKHW-----EFSSLRRSK 1686
QY 1345 WPLRCKLQORTLYELHYQLITFGK---VFCTKSRNACNACPMRGECRHFASAYASARLAL 1401
Db 1687 WSTLC-----MLVELHQ-----GODRFVY-----TCNECKHHVETRWHTCVCEYDLCI 1731
QY 1402 PAPERSLTSATIPVPSPFPVAPIMIELPLPL-EKSLASGAPSNRENCEPIIEEPASP 1460
Db 1732 NCYNTKSHTH-----KMKWGLGLDDEGSSQGEPOSK-----SP 1765
QY 1461 GOCEITEIT-----ESDIEDAYYNEDPEITIKLNTIEOGMTLREHMERNMELQEG--- 1511
Db 1766 -QESRRLSTQRCIQSLIVHACQCRNANCSLPSC-----QMKRVVQHTKCKRK 1812
QY 1512 -----DMSKALVAL-----HPTTSITFP---KUK-NISRLRTEHQVYVELPDSHRLD 1556
Db 1813 TNGGCPVCKQLLALCCYHAKHCQENKCPVPFCLNKHNVROQOIQHCLQQAOLMRRMAT 1872
QY 1557 MDKR---BPDDPSYLLAIWTPE---TANSAPPEQ 1587
Db 1873 MNRNVPOQSLSPTSAPPTQPTQPTQPPAQ 1909

RESULT 10

US-08-598-591-2
; Sequence 2, Application US/08598591
; Patent No. 5654155

GENERAL INFORMATION:

; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Murphy Patricia D.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Scheiter, Denise B.
; APPLICANT: Zeng, Bin

; TITLE OF INVENTION: A Consensus Sequence of the Human BRCA1 Gene

; Patent No. 5654155

; NUMBER OF SEQUENCES: 74

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince St.
; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22314

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release
; CURRENT APPLICATION DATA:
; - FILING DATE: herewith
; APPLICATION NUMBER: US/08/598.591

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Swecker, Robert S.

; REGISTRATION NUMBER: 19,885
; REFERENCE/DOCKET NUMBER: 020160-282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME: 17
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
US-08-598-591-2

Query Match

Best Local Similarity 2.0%; Score 182; DB 1; Length 1863;

Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

QY 20 DGSRODVLEF---DLNKTQQKPKRRKFKMPKVVVVEGPKR----- 58
Db 219 DSAKKAACEFSDVTNTEHQPSNNDLNTTEKRAAERHPEKYQGVSSVSNLHVPCGTNT 278
QY 59 -----KPKPAELPKVVVEGPK-----KKPKKAATOEVKSKETSACK 98
Db 279 HASSLOHENSLLLTQDRNMNVEKAEFCNKSQKOPGLARSOHNWAGSKETCNDRRTPTSEK 338
QY 99 K-----NLESATKKPANYGDMNSKSPVTLKSCRKALN-----FDLEN 137
Db 339 KYDLNADPLCKERKNKOKLPCSENPROTEV---PWITLASSIQKYNWFVSRSDLEAG 394
QY 138 PGDARQGDSESEI-----VONSSGANS-----FS 161
Db 395 SDDSHDGESENAKADVLDVLNVEYDSEKIDLLASDPHEALICKSERVHKSYES 454
QY 162 EIRDAIGGTN-----GSFLDSVSQIDK---TNGLGAMNOPLVSMG 199
Db 455 NIEDKIFGKYTKKASLPNLHVTEENLIIGAVTPEIIQERPLTNKLRKRRP---TSG 511
QY 200 NOPDKLSTGAKLARDQPDLLTR---NOQCQFPVATQNTQFPMENOQAWLQMKNLIGPFP 257
Db 512 LHPEDFIKKADLAVOKTPEMINQGTQEQNGQVMNITNSGHENKTGDSIQNE----- 565
QY 258 GNOQPRWIRNOOPCLAMGNOOPMYLIGTPRPALVSGNOQLGGPOGNKRPIFLN-HQTCL 316
Db 566 -----KNPNPIESLEKESAFKTAEPISISSIS-NMELE-----LNTHNSKA 605
QY 317 PAGNOLYSGPTDMH---OLVMS-----TGGQOHGLLILKNOOPGSLIRG 356
Db 606 PKKNRLRRKSSRTHIHALELVVSRNLSPNPNCTELQIDSCSSSEETKKKYNOMP---VRH 662
QY 357 QQPCVPLIDQOPAT-----PKGFT-----HLNQMV 381
Db 663 SRNLQIMECKEPATGAKKSNKPEQTSKRHSDTTPPELKLITNAPGSFTKCSNTSELKEFV 722
QY 382 ATSM-----SSPGLRPHSQSOVP---TTLHVES 407
Db 723 NPSLPREKEEKELETVKVSNNAEDPKDMLSGERVLTQERSVSSISILVPGDTGYTQES 782
QY 408 VSRILNGTGTGTCOR-----SRAPAYDSIQODIH-----OGNKYILSHEISNGNG 451
Db 783 ISLLEVSTLIGKAKTEPNKCVSQCAAFENPKGLIHGCKDNDRNDTEGFKYPLGHEV----- 837
QY 452 CKKALPQNSSLPTIMAKLEEA-----RGSKRO-----YHRA 483
Db 838 -----NHSRETSI--EMESELDAQYLONTFTFKYSKRSQSFALFSPNPGNAEECATFSAH 888

```

484 MGQTEKHLNLAQIQSQDVERHNSSTCEVYLDAAKTKTIQKVQENL-HGMPPEVIEI 542
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
889 SGLKQSPKVPTECEQKEENQGNES-----NKPQVTNITAGFP--VVGQ 934
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
543 EDDPTDAR---GKNTASIS---KGASKNSPVPKTAKEKCIYP-----KTPA 587
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
935 KDKPVDNAKSGIRKGSFRCLSSQFRNETGLITPNKHGLLQNPYRIPPLPFIKSFVTKC 994
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
588 KKGKGRKKSVPVPAHASELQWQTPPKT--PLSRSKP-----KKGKRSIODSGKARG 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
995 KKNL---EENFEHMSPEEMGNENIPSTVTSIRNIRENVFKGASSNINEVGSSTN 1052
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
641 PSGELLQDSTAEIIRYQNL---LGBKEREQONAMVLYKGDALVPYESKK----- 691
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1053 EVG-----SINEIGSSDENIAELG--RNRGPKLNAMRL--GVLOPEYVKSLPSCSNC 1103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
692 KRPKPKVIDDETTRIWL-----LMRGKDEKEGD 721
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1104 KHPEIKKQEEYEVQVTNTDFSPYLSIDNLEQPMGSSHASQVCSETPDDLDDGEIKEDT 1163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
722 E--EKDKKKKEWEEERRV---RG---RADSFIAHMLVQDREFFSPWKGSVVDSVI 771
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1164 SFAENDIK-----ESAVFSKSVQRELSRSPFTHHLAQYRGA----- 1206
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
772 GVFLTONVSDHLSSAPMSLAARFPFKLSSRREDERNVRSVVVWEDPE--GCILML-----N 825
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1207 -----KLESS-----EENLSS---EDELPCFQHLFGKVN 1235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
826 EIPSWQKVQHPDSMEVSGVDSGSKE-----OLRDCNSGIERNFLEKSTONLEEV 878
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1236 NIPS--QSTRH--STVAPECLSKNTEENLLSKNSLNCDSNOVI-----LAKASQ---EHH 1284
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
879 LSSQDSFDPALFQSCGRVSGSCSDAEPTTRCTETKTVSGTSOSVQGTSPN---LSD 934
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1285 LSBEKCSALFSS-----QSELEDLTANTQDPFLIGSKOMRHOSESQGVCLSD 1337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
935 EICLQGNRPHLYEGSDVQKQETTNVAQKPDLEKTMNMKDSVCFQGPQNDTNWQTPS 994
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1338 KELVSDDEE---RGTG---LEENN--QEEQMSDNLGEAASGC-----ESET 1376
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
995 SSVYOCATROPHVLDIEDFGMQEGGLGYSNWSISPRVDRVKNKNVPRFRFQSGSVPRF 1054
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1377 SVSEDCS----- 1383
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1055 TGOIIPSTHELPMGLGSSSAVQEHQDPT--QHNQOENMKASHLQKTFDLNLSSEEC 1113
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1384 -----GLSSQSDILTTQORTMQHN-----LIKLOQEMAELEAVLEQH 1421
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1114 LTROSSTKQNTDGCPLPRDRFTAEDVDPLSNSSQLNVLVSNSSKNKEQTAVEYKETNAT 1173
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1422 GSQPSNSYPSI-----ISDSSALEDLRNPQESTSEKAVLTSSOKSSEYP 1464
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1174 ILREMGKTLADGKKPTQOWDSLKRDVEGNGRGRNKNMDSIDYEAIRASISEISEAI 1233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1465 ISONPEGLSAD--KFEVSADSSSTKNKE--PGVERSPSKPCPSLD-----DRWMHSCSGSL 1517
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1234 KERGM--NMMLAVRIKDFLERIVKDHGGIDL 1262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1518 QNRNYSQEEELIKVVDVEEQLEESGPHDL 1547
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-08-798-691-2
; Sequence 2, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
```

```

; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
;
US-08-798-691-2
```

Query Match 2.0%; Score 182; DB 1; Length 1863;

Best Local Similarity 18.4%; Pred. No. 8.9e-06;

Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

```

QY 20 DGSRDQVLEF---DLNKTPOQKPSKRRKPKMPKVVEGPKR----- 58
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 219 DSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPKRYOGSSVSNLHVPCGTNT 278
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 59 -----KPRKPAELPKVVVEGK-----KRPKKAATQEKVKKSGTSAKK 98
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 279 HASSLOHENSLLLTKDRMNVKAEFCNKSQKPLARSOHNRWAGSKETCDRTPTSTEK 338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 99 K-----NLKESATTKPANVGMDSNKGPEVTLKSCRKALN-----FDLEN 137
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 339 KVDLNADPLCKERKWNKQLPCSENPRDTEV---PWITLSSIQKVNWFNFSRDELLG 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 138 PGDARQGDSESEI-----VQNSSGANS-----FS 161
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 395 SDDSHDGESESNAKVADVLDVLNEVDEYSGSEKIDLLASDPHEALICKSERVHSKSVES 454
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 EIRDAIGGTN-----GSFLDSVSQIDK---TNGLGAMNOPLEVSMG 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 455 NIEDKIFGKYRKKASLPNLSHVNTENLLIGAFVTEQIIQERPLTNKLRKRP---TSG 511
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 NOPDKLSTGAKLARDQOQDILLTR--NQOQFVATQNTOPFMENQQAOLMKQLIGFPF 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 512 LHPEDEFIKRADLAVQKTPPEMINQGTQEQNGQVMNITNSGHENKTKGDSIQNE----- 565
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 258 GNOQPRMTIRNQOPLCAGMGNOQPMYILGTPRALVSGNOOLGCGPOGNKRPIFLN--HOTCL 316
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 566 -----KNPNPIESLEKESAFKTAEPISISSI--NMELE-----LNIHNSKA 605
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```


Db 279 HASSLOHENSLLLTDRMNVKAEFCNCKSQPLARSOHNRWAGSKETCDNRRTSTPEK 338
 QY 99 K-----NLKESATKKPANVCDMSNKSPEVTLKSCRALN-----FDLEN 137
 Db 339 KVDLNADPLCEKWNKQKLPSENPRDTEV-----PWITLNSIQKVNWFSSRDELIG 394
 QY 138 PGDARQDSESEI-----VONSSGANS-----FS 161
 Db 395 SDSHOGESSENAKADVLDVLENEVEYSGSSEKIDLLASDPHEALICKSERVHSKSVES 454
 QY 162 EIRDAIGGTN-----GSFLDSVSQIDK-----TNGLGAMNPLEVSMG 199
 Db 455 NIEDKIFGTYRKASILNPSLHVNTENLIIGAFVTEQIIQOERPLTNKLRKRPP-----TSG 511
 QY 200 NOPDKLSTGAKLARDQOPLLTR--NOOCOPVATONTQPMENQOAWLOMKQNLIGFPF 257
 Db 512 LHPEDFIKKADLAVKTPENINGTQTEONGOVMINITNSGHENKTKGDSIQNE-----565
 QY 258 GNOOPRMTIRNOOPCLAMGNQOQPMYLGTPRALVSGNOOLGPGQGNKRPILFN--HOTCL 316
 Db 566 -----KNPNPIESLEKESAFKTAEPISISSIS--NMELE-----LNIHNSKA 605
 QY 317 PAGNQLYGSTDMH-----QLVMS-----TGGQOHLIKNOQPGSLIRG 356
 Db 606 PKNLRKRSSTRIHALELVWNLSPNCTELQIDSCSSSEIKKKYNOMP-----VRH 662
 QY 357 QOPCVPLIDOPAT-----PKGFT-----HLNQMV 381
 Db 663 SRNLQMEGKEPATGAKSKNPNQTSKRHSDTFPELKLITNAPGFTKCSNTSELKEFV 722
 QY 382 ATSM-----SSPLRPHSQSQVP--TYTLHVES 407
 Db 723 NPSLPREEKEKLETVKVSNNAEDPKDLMLSGERVLTQERSVSSISLVPDGYGTQES 782
 QY 408 VSRILNGTTCOR-----SRAPAVDSLOODIH-----QGNKYLSHSINGNG 451
 Db 783 ISLEVSTLGAKTEPNKCVSQAAPENPGLIHGCSKDNRNDETEGFKYPLGHEV-----837
 QY 452 CKKALPONSLPTPIAKLEEA-----RGSKRQ-----YHRA 483
 Db 838 -----NHSRETSI--EMESELDAQYLONTFKVSKRQSFALFSPGNAEECATFSAH 898
 QY 484 MGTEKHDLNLAQIAQSDQVERHNSSTVEYLDAAKTKIKVQOENL--HGMPPEVIEI 542
 Db 889 SGSLKQSPKVTCEQKEENOCKNES-----NIKPVQTVNITAGFP--VVGQ 934
 QY 543 EDDPTDGR--KGKNTASIS--KGASKNSSPVKTAKEKCIVP-----KTPA 587
 Db 935 KKPVDNAKSIKGGSRFCUSSQFRNETGLITPNKHGLLONPYRPLPPIKSFVKTKC 994
 QY 588 KKGKGRKKSVPVPPAHASEIQLWQTPPKT--PLSRSKP-----KKGKRSIQDSGKARG 640
 Db 995 KKNLL--EENFEHSMSPEREMGENIPSVTISRNIENFVKGASSNINEVGSSTN 1052
 QY 641 PSELGQDSIAEIIYRMQNL--LGDKEREQONAMVLYKGDGALVPEYSK-----691
 Db 1053 EVG-----SSINEIGSDENIQALG--RNRGPKLNLMLR--GLVQPEYKQSLPGSNC 1103
 QY 692 RKPRPKVDIDETTRIWNL-----LMGKGDEKEDG 721
 Db 1104 KHPETIKQEEYVVQVNTDFSPYLSIDNLEQPMGSSHASQVSETPDDLLDDGEIKEDT 1163
 QY 722 E--EKDKKKEKWEERRVF--RG--RADSFIAHMLVQDGRFRSPWKGSVVDSVI 771
 Db 1164 SFAENDIK-----ESSAVFESKSVQRCGLSRSPSPTHTHLAGYRGA-----1206
 QY 772 GVFLTONVHLSSAFMSLAARPPKLSRREDENRVSVVWDEP--GCILNL-----N 825
 Db 1207 -----KKLESS-----EENLSS--EDELPCFQHLHFGKVN 1235
 QY 826 EIPSWOEKVOHPDMEVSGVDSKSE-----QLRDCNSNGIERNFLEKSTONLEEEV 878
 Db 1236 NIPS--OSTRH--STVATECLUSKNTNENLJSLKNSLNDSCSNQVI-----LAKASQ-----EHH 1284

QY 879 LSSQDSFDPALFOSCGRVGSCSKSDAEPTTRCTETKTVSGTSQSVQTCSPN----LSD 934
 Db 1285 LSEETKCSASLFS-----QCSELEDITANTNTODPFLIGSKQMRHOSQGVCLSD 1337
 QY 935 EICLOGNERPHLYEGSDGVQKQETTNVAQKKPDLKTMNMKDSVCFGQPRNDTNWQTTPS 994
 Db 1338 KELVSDDEE--RGTG--LEENN--QEEQMSDNLGEAASGC-----ESET 1376
 QY 995 SSVTEQCATROPHVLDIEDFCMQGEGGLYSWMSISPRVDRVKNKNVPRFRFQSGSVPRF 1054
 Db 1377 SVSEDCS-----1383
 QY 1055 TGOIIPSTHELPCMGLSGSSSAVQEHODDT--OHNOQDEMKNKASHLOKTFDLNLSSEEC 1113
 Db 1384 -----GLSSQSDILITQORTMQHN-----LIKLOQEMAELEAVLEQH 1421
 QY 1114 LTRQSTKONITGCLPRDRTAEDVVDPLSNSSLONLVSNSSNKEQTAVEYKETNAT 1173
 Db 1422 GSQPSNSYPSI-----ISDSSALEDLRNPQEQSTSEKAVLTSQKSSEYP 1464
 QY 1174 ILEMKGTLADGKKPTQOWDSLRKDYEGNEGROERNKNMDSIDYEAIRRASISEISEAI 1233
 Db 1465 ISQNPEGLSAD--RFEVSADSSSTSKNKE--PGVERSSPKCPSLD----DRWYMHSCSGSL 1517
 QY 1234 KERGM--NNMLAVRIKDFLERIVKHGIGIDL 1262
 Db 1518 QNRYPSQBELIKVVDVEEQLEESGPHDL 1547
 RESULT 13
 US-08-825-487A-2
 : Sequence 2, Application US/08825487A
 : Patent No. 6048689
 : GENERAL INFORMATION:
 : APPLICANT: Murphy, Patricia D.
 : APPLICANT: White, Marga B.
 : TITLE OF INVENTION: METHODS FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE
 : NUMBER OF SEQUENCES: 110
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Howrey & Simon
 : STREET: 1299 Pennsylvania Avenue., N.W.
 : CITY: Washington,
 : STATE: DC
 : COUNTRY: USA
 : ZIP: 20004
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/825,487A
 : FILING DATE: 28-MAR-1997
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US98/060002
 : FILING DATE: 26-Mar-1998
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Albert P. Halluin
 : REGISTRATION NUMBER: 25,227
 : REFERENCE/DOCKET NUMBER: 05371.0012.999
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 650-463-8100
 : TELEFAX: 650-463-8400
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1863 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: not relevant
 : TOPOLOGY: not relevant
 : MOLECULE TYPE: protein

ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCAL
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-08-825-487A-2

Query Match 2.0%; Score 182; DB 3; Length 1863;
Best Local Similarity 18.4%; Pred. No. 8.9e-06;
Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

QY 20 DCSRDVLEF---DLNKTPOQPSRKRKFMKVVVEGPKR-----58
DB 219 DSAKKAACEFSDTNTTEHHQPSNDLMTTEKRAAERHPKYQGSVSNLHVEPCGTNT 278
QY 59 -----KPRKPAELPVVVEGKP-----KRPKKAATOEVKSKETGSACK 98
DB 279 HASSLOHENSLLLTDRMNVEKAFCNKSQOPGLARSOHNWAGSKETCDRRTPSTEK 338
QY 99 K-----NLKESATKPKPANYGDMNSKSPVTLKSCRKALN-----FDLEN 137
DB 339 KYDLNADPLCERKENWKQKLPCESENPRTEDEV---PWITLASSIQKYNWESRSDDELLG 394
QY 138 PDARQGDSESEI-----VONSSGANS-----FS 161
DB 395 SDDSHDGESESNKADVLDVLNEVDEYSGSEKIDLLASDPHEALICKSERVHSKSVEB 454
QY 162 EIRDAIGTN-----GSFLDSVSQIDK---TNGLAMNQPLEVSG 199
DB 455 NEDKIFGTYTKKASLPNLSHVTENLIIGAFVTEPQIIIOERPLTNKLRKRPP---TSG 511
QY 200 NOPDKLSTGAKLARDQPDLLTR--NQOCQFPVATQNTQPMENOQAMLOMKNQLIGPFP 257
DB 512 LHPEDFIKKADLAVQKTPMINQNTQTEQONQVNNITNSGHENKTGDSIGNE-----565
QY 258 GNOQPRWIRNOOPLANGNOOPMYLIGTPRALVYSGNOOLGGPOGNKRPIFLN--HOTCL 316
DB 566 -----KNPNIESLEKESAFKTAKEPISSIS--NMELE-----LNHNSKA 605
QY 317 PAGNOLYSGPTDMH---OLVMS-----TGQOQHGLLKNQOQSGSLRG 356
DB 606 PKNLKRKSTTRHIALELVVSRNLSPPNCTELOIDSCSSSEIEKKKYNQMP---VRH 662
QY 357 QQPCVPLIDQOPAT-----PKGFT-----HLMQAV 381
DB 663 SRNLQMEGKEPATGAKKSNKPNQTSKRHSDTTPPELKLTPAGSFTKCSNTSELKEV 722
QY 382 ATSM-----SSPGLRPHSQSQVP--TTYLHVES 407
DB 723 NPSLPREEKEELETVKVSNNAEDPKDMLSGERVLTQTERSVESSISLVPCTDYGTQBS 782
QY 408 VSRILNGTGTGTCOR-----SRAPAYDSLOODIH-----QGNKYILSHETSNGNG 451
DB 783 ISLEEVSTLGKAKTBNKCVSOCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEV-----837
QY 452 CKKALPQNSLPTPIAKLEEA-----RGSKRO-----YHRA 483
DB 838 -----NHSRETSI--EMEESLDAQYLQNTFKVSKRQSFALSNPGNAEBECATFSAH 888
QY 484 MQQTEKHLNLAQQAQSOQDVERHNSSTCVYLDAAKTKIKQVVOENL--HGMPPVEIEI 542
DB 889 SGSLAKQSPKVTFECEQKBENOGKES-----NIKPVQTVNITAGFP--VVQG 934
QY 543 EDDPTDGR---KGNTASIS---KGAGNSSPVKKTAEKECIVP-----KTPA 587
DB 935 KDKVDNKAISKSGSRFCSSQFRGTEGLTIPNKHGLLQNPYRIPPIPKSFVKTKC 994
QY -588 KKGAGRKSVPPPAHASEIOLWQTPPKT---PLSRSKP-----KGKGRKSQDQSGKARG 640
DB 995 KKNLL--BENFEHNSPSPREMGNEINIPSTVSTISRNNIRENVFKGASSNINEVGSSTN 1052

QY 641 PSCELLCQDSIAEILYRMONLY--LGDKEREQEQNANWLYKGGALVPYESKK-----691
DB 1053 EVG-----SSINEIGSSDNIQAEIG--RNRPKLNLAMRL---GVLOPEVYKOSLPGSNC 1103
QY 692 RKPRKVIDDDETTIRIWL-----LMGKGDEKEGD 721
DB 1104 KHEIKKQYEEVQTVNTDFSPYLLISDNLEQPMGSSHASQVCSETPDLDLDDGEIKEDT 1163
QY 722 E--EKDKKKEKWEERRVF-----RG-----RADSFARMHLYOGDRRFRPMKGVVDSVI 771
DB 1164 SFAENDIK-----ESSAVFSKSVQGELSRSPSPFTHLAQGYRGA-----1206
QY 772 GYFLQNTQNSDHLSSAFNSLAARFPKLLSSSRREDERNVSVVVEDEPE--GCILNL-----N 825
DB 1207 -----KKLESS---EENLSS-----EDELPCFOHLLFGKYN 1235
QY 826 EIPSWQEKVOHPDMEVSGVDSGSKE-----QLRDCSNSGIERNFLEKSIQNLEEV 878
DB 1236 NTPS--QSTRH--STVATECLSKNTEENLLSKNSLNDCSNOVI-----LAKASQ---EHH 1284
QY 879 LSSQSFDPAPFQSCGRVSGSCSKSDAEFPTRTCTKTVSGTSOSVOTGSPN---LSD 934
DB 1285 LSEETKCSASLFSS-----QCSELEDLTANTQDPFLIGSSKOMRHOSESQGVGLSD 1337
QY 935 EICLOGNERPHLYEGSDVQKQETTINVAQKPDLEKTMNWKDSVCEGQPRNDTNWQTFS 994
DB 1338 KELVSDDEB---RGTG---LEENN--QEEQSDMSNLGEAASGC-----ESET 1376
QY 995 SSYEOCATRQPHVLDIEDFGMOGEGLYGWSMISPRVDRVKNKNVPRFRFGGSPREF 1054
DB 1377 SVSEDCS-----1383
QY 1055 TQGIIPSTPHELPGMLSGSSSAVOEHQDDT--OHNOODEMNKASHLQKTFDLNLSSEC 1113
DB 1384 -----GLSQSDILITQQRDTMOHN-----LIKLOEAMAELEAVLEQH 1421
QY 1114 LTRQSTKQNIYDGLPRDRTAEDVVDPLSNSSNLQNLIVESNNSKEQTAVEYKETA 1173
DB 1422 GSQPSNSYPSI-----ISDSALEDLRNPQSTSEKAVLTQSXSSEYP 1464
QY 1174 ILREMGKTLADCKKPTQOWDSLKRDVEGNEGQERKNMNDSDYEAIRRASTSEIAI 1233
DB 1465 ISQNEGLSAD--KFEVADSSTSKNKE--PGVERSPSKCPSLD-----DRWYHSCGSL 1517
QY 1234 KERGM--NNLAVRIKDFLERIVKDHGGIDL 1262
DB 1518 QNRNYPQSEELIKVYDVEEQLEESGPHDL 1547

RESULT 14
US-08-825-487A-6
; Sequence 6, Application US/08825487A
; Patent No. 6048689
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: White, Marga B.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue., N.W.
; CITY: Washington,
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,487A
; FILING DATE: 28-MAR-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA: PCT/US98/060002
FILING DATE: 26-Mar-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 05371.0012.999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8100
TELEFAX: 650-463-8400
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCA1
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-08-825-487A-6

Query Match 2.0%; Score 182; DB 3; Length 1863;
Best Local Similarity 18.4%; Pred. No. 8.9e-06;
Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;
QY 20 DGSQDVFLEF---DLNTPQOKPSKRRKFKMPKVVVGKPKR-----58
DB 219 DSAAKACAESETDVTTEHQPPNSNDLNTTEKRAAERHPEKYGGSSVNLHVEPCGTNT 278
QY 59 -----KPRKPAELPKVVVEGKP-----KRPKRAATOPKVKSKETGSAKK 98
DB 279 HASSIQHENSLLLTDKRMNVKAEFCNKSKQPGKLARSQHNWAGSRETCDNRTPSTEK 338
QY 99 K-----NLKESATKKPANVGMNSKSPVTLKSCRKALN-----FDLEN 137
DB 339 KVDLNADPLCKERKWNKQKLPCESENPDTEV-----PWITLNSIQKVNWFNSRDELLG 394
QY 138 PGDARQDSESEI-----VQNSSGANS-----FS 161
DB 395 SDDSHGDESESNKAVADVLDVNEVDEYSGSEKIDLLASDPHEALICKSERVHSKSVES 454
QY 162 EIRDAIGTN-----GSFLDSVSQIDK-----TNGLGAMNOPLEVS MG 199
DB 455 NIEDKIFGKTYRKASLPNLSHVNTENLIIGAFVTEPQIIQERPLTNKLRKRRP---TSG 511
QY 200 NOPDKLSTGAKLARDQPDLLTR--NQOCQFPVATQNTQPMENOQAWLQKWLIGFPF 257
DB 512 LHPEDFTFKADLAVQKTPENINQNTQTEQGVMMNTNSCHENKTKGDSIQNE-----565
QY 258 GNOQPMRTIRNOOCLAMGQOQPMYLLIGTPALVSNQOOLGGPGQNKRFIEFN-HOTCL 316
DB 566 -----KNPNPIESLEKSAFKAEPISSIS-NMELE-----LNHNSKA 605
QY 317 PAGNQLGSPDMDH-----QLVMS-----TGGOQHGLLIKNOQPGSLIRG 356
DB 606 PKNRLRRKSSTRHIALELVSNLSPNPPNCTELQIDSCSSSEIEKKKYNQMP---VRH 662
QY 357 QQPCVPLIDQOPAT-----PKGFT-----HLNQMV 381
DB 663 SRNLQMEGKEPATGAKSKNPNQTSKRHDSDTFPPELKLTPNAPGSFTKCSNTSELKEFV 722
QY 382 ATSM-----SSPGLRPHSQSOVP-TTYLHVES 407
DB 723 NPSPREKEKEKLETVKVSNNAEDPKDMLSGRVLQTERSSESSISLVPGTDYGTQES 782
QY 408 VSRILNGTTCTCOR-----SRAPAYDSLOODIH-----QGNKYILSHEISNGNG 451

DB 783 ISLEVSTLGAKEPKNKCVSOCAAFENPKGLIHGCKSDNRNDTEGFKYPLGHEV-----837
QY 452 CKKALPONSSLPPIMAKLEEA-----RGSRRQ-----YHRA 483
DB 838 -----NHSRETSI---EMBESELDQYLONTFKVSKRQSFALFSPGNAAEEECATSAH 888
QY 484 MGOTEKHDNLAAQIAQSDQVERHNSSTCEVYLDAAKTKTIQKVVOENL-HGMPPEVIEI 542
DB 889 SGLKQSPKVTPECEKEENOGKNES-----NPKPVQVNIITAGFP--VVGQ 934
QY 543 EDDPTDGR---KGKNTASIS---KGASKNSSPVKTAKEKCIVP-----KTPA 587
DB 935 KDKPVDNAKSGIKGSRFCLSQFRNETGLITPKHGLLQNPYRIPPLFPKSFVTKC 994
QY 588 KKGRAGRKKSVPPPAHASEIOLWQTPPKT--PLSRSKP-----KGKGRKSIODSGKARG 640
DB 995 KKNLL--EENFEHMSPEREMENIPSTVSTISRNIENRVFKGASSNINEVGSGSTN 1052
QY 641 PSGELLQODSIAEIIYRMQNLV--LGDKEREQONAMVLXKGDGALVPYESKK-----691
DB 1053 EVG-----SSINEIGSSDENIQAELG--RNRGPKLNAMRL--GVLQPEVYKQSLPGSNC 1103
QY 692 KPRPKVDIIDEETRIWNL-----LMKGDEKEGD 721
DB 1104 KHPEIKKQEEYEVVQTVNTDFSPYLIISDNLEQPMGSSHASQVCSPTDLDLDDGEIKEDT 1163
QY 722 E--EKDKKKKEKWEERRVF---RG---RADSFIAHMLVQGDRRFSPWKGSVVDVSI 771
DB 1164 SPFENDIK-----ESSAVFSKSVORGELSRSPFPFTHLAQYRGA-----1206
QY 772 GVFLTONVSDHLSSAPMSLAARFPFKLSSRREDERNVRVSVVDEPPE-GCILNL-----N 825
DB 1207 -----KKLESS---EENLSS---EDELPCFQHLFLFKVN 1235
QY 826 EIRSWQEKVQHPDSMEVSGVDSSKE-----QLRDCSNSGIERFNFLEKSTONLEEV 878
DB 1236 NIPS--OSTRH-STVAPECLSKNTEENLLSKNSLDCSNQVI-----LAKASO---EHH 1284
QY 879 LSSQDSFDPALFQSGRGVGSCSKSDAEFPPTTTCETKTVSGTSQSQVQTSN---LSD 934
DB 1285 LSEETKCSALFSS-----QCSELEDLTANTQOPFLIGSKQMRHQSOGVGLSD 1337
QY 935 EICLQNGRPHLYEGSGDVQKQETTNVAKKPDLEKTMNMKDSYCFQGPQRNDTNWQTTPS 994
DB 1338 KELVSDDEE---RGTG---LEENN--QEEQSMDSNLGEAASGC-----ESET 1376
QY 995 SSTEQCATRPHVLIDIEDFCMOGEGLYSWMSSISPRVDRVKNKNVPRFRFQGGSVVPREF 1054
DB 1377 SVSEDCS-----1383
QY 1055 TGOIIPSTPHELPGMGLSGSSSAVQEHODDT-OHNQODENMKASHLOKTFDLDLNSSEEC 1113
DB 1384 -----GLSSQSDILTTQORDTMQHN-----LIKQOEMAELEDAVEJQH 1421
QY 1114 LTRQSTKQNTDGLCPDRDRTAEDVVDPLSNSSLNILVESNSSNKEQTAVEYKETNAT 1173
DB 1422 GSOPNSYPSI-----ISDSSALEDLRNPQESTSEKAVLTQKSSSEYP 1464
QY 1174 ILREMGKTLADGKKPTQOWDSLRKDVNEGROERNKNNMDSIDYEAIRASISEISEAI 1233
DB 1465 ISONPEGLSAD-RFEVSADSTSKNKE--PGVERSSPKCPSLD-----DRWMHSCSGSL 1517
QY 1234 KERGM--NMMLAVRIKDFLERIVKDHGGIDL 1262
DB 1518 QNRNYPQSEELIKVVDVEEQOLESGPHDL 1547

RESULT 15
US-09-074-476-2
; Sequence 2, Application US/09074476
; Patent No. 6130322
; GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.
APPLICANT: Allen, Antonette C.
APPLICANT: Alvares, Christopher P.
APPLICANT: Critz, Brenda S.
APPLICANT: Olson, Sheri J.
APPLICANT: Thurber, Denise
APPLICANT: Zeng, Bin
TITLE OF INVENTION: Coding Sequences of the Human
TITLE OF INVENTION: BRCA1 Gene
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue N. W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074.476
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/074.453
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 5371.34.US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8109
TELEFAX: 650-463-8400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCA1 (om11)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-09-074-476-2

Query Match 2.0%; Score 182; DB 4; Length 1863;

Best Local Similarity 18.4%; Pred. No. 8.9e-06;

Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

QY 20 DGSRODVLFE---DLNKTPOOKPSKRKRKFKPVVVEGPKR-----58
DB 219 DSAKKAACEFSETDVTNTEHHQPSNDLNTTEKRAAERHPKYQGSVSNLHVPCGTNT 278
QY 59 -----KPKRPAELPKVVVEGKP-----KPKRKAATQEKVKSKETGSACK 98
DB 279 -HASSLOHENSLLLTDRMNVKAEFCNKSQKQPLARSOHNRWAGSKETCDNRTPSTEK 338
QY 99 K-----NLKESATKPKANVGMNSKSPVETLKSCRKALN-----FDLEN 137
DB 339 KVDLADPLCERKENWKQKLPCCSENPRTEDEV-----PWITLNSSIQKYNWESRDELLG 394
QY 138 PCDAQQDSESEI-----VQNSSGANS-----FS 161
DB 395 SDDSHDGESNAKVADVLDVINEYDEYSGSEKIDLLASDPHEALICKSERVHSKSVES 454
QY 162 EIRDAIGTGN-----GSPFSDVSQIDK---TNGLGANNQPLEVSG 199

DB 455 NIEDKIFGKYRKASLPNLSHVTEENLIIGAFVTPBPIIOERPLTNKLRKRRP-----TSG 511
QY 200 NOPDKLSTGAKLARDQPDLLTR--NOOCQFPVATQNTQFPMENOQAWLQMLGIFPPF 257
DB 512 LHPEDFIKKADLAVQKTPEMINQGTQTEQNGQVMNITNSGHENKTGDSIQNE-----565
QY 258 GNOQPRMTIRNOOPCLAMGNQOPWYLIPTPRALYSGNQOGLGPGQGNKRPIFLN-HOTCL 316
DB 566 -----KNPNPIESLEKESAFKTKAETPSSSSIS-NMELE-----LNTHNSKA 605
QY 317 PAGNOLYGSPTDMH-----QLVMS-----TGQOQHGLLIKNOQPGSLIRG 356
DB 606 PKNRLRRKSSSTRHIHALELVVSRNLSPPNCTELOIDSCSSSEEEKKKYNQMP---VRH 662
QY 357 QOPCVPLIDQOPAT-----PKGFT-----HLNOMV 381
DB 663 SRNLQLMGKKEPATGAKKSNKPNQTSKRHSDTTPPELKLTNAPGSFTKCSNTSELKEFV 722
QY 382 ATSM-----SSPGLRPHSQSOVP--TTLVHVES 407
DB 723 NPSLPREEKEKLETVKVSNNAEDPKDMLSGERVLOTERSSESSISLVPDGTGTQBS 782
QY 408 VSRILNGTTGTQCR-----SRAPAYDSLQDDIH-----QGNKYILSHEISNGNG 451
DB 783 ISLLEVSTLGRKAKTEPNKCVSQCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEV-----837
QY 452 CKKALPONSSLPPTIMAKLEEA-----RQSKRQ-----YHRA 483
DB 838 -----NHSRETSI--EMEESLDAQYLQNTFKVSKRQSFALFSPNPGNAEECATFSAH 888
QY 484 MGQTEKHDNLNLAQOJAQSDQDVERHNSSTCEVYLDAAKTKIKQVVQENL-HCMPPEVIEI 542
DB 889 SGLSKKQSPKVTFECEQKEENOGKNES-----NIRPVQTVNITAGFP--VVGQ 934
QY 543 EDDPTDGRAR--KGKNTASIS--KGASKGNSSPVKKTAEKEKCIVP-----KTPA 587
DB 935 KDKPVDNAKCSIKGSRCLSSQFRNETGLITPNKHGLLQNPYRIPPLPIKPSFKVTKC 994
QY 588 KKGRAGRKSVPPPAHASEIQLWQTPPKT--PLRSRKP-----KGKGRKSTODSGKARG 640
DB 995 KKNLL--BENFEHSMSPEREMGNENIPSTVTSIRNNIRENVNFKGASSNINEVGSSTN 1052
QY 641 PSELLCQDSIAEIIYRMONLY--LGDKEREQOANWLYKGDGALVPVYESK-----691
DB 1053 EVG-----SSINEIGSSDENIOAELG--RNRGPKLAAMLRL---GVLPQEVYKQSLPGSNC 1103
QY 692 RKPRPKVDIDDETTIRIWNL-----RADSFIAARMHLVQDGRFRFPWKGSVVDSVI 721
DB 1104 KHPEIKKQYEVEVQTVNTDFSPYLISDNLEQPMGSSHASQVCSFETPDLLDDGEIKEDT 1163
QY 722 E--EKDKKKEKMWEEERRVF-----RG-----RADSFIAARMHLVQDGRFRFPWKGSVVDSVI 771
DB 1164 SPAENDIK-----ESSAVFSKSVQSGELSRSPFTTHLAQGYRGA-----1206
QY 772 GVFLTQNVSDHLSSSAFMSLAARFPFKLSSSRREDNRVSVVVEDEPE-GCIINL---N 825
DB 1207 -----KKLESS---EENLSS-----EDELPCFPHLLFGKVN 1235
QY 826 EIPSWQERKVOHPDMEVSGVDSGSKE-----QLRDCSNSGIERNFLEKSIQNLLEEV 878
DB 1236 NIPS--QSTRH--STVATECLSKNTEENLLSKNSLNDCSNQVI---LAKASQ---EHH 1284
QY 879 LSSQDSFDPAPQSCGRVSGSCSKSDAEFFPTTRETCTVTSQTSOSVQGSFN-----LSD 934
DB 1285 LSEETKCSASLFS-----QCSELEDITANTQDPFLIGSSKQMRHQSESOGVGLSD 1337
QY 935 EICLOGNERPHLYEGSGDVQKQETTNAQKPDLEKTNWKNKDSVCFQGPDRNDTNTQTPS 994
DB 1338 KELVSDDEE---RGTG---LEENN--OEQOSMDNLGEAASGC-----ESET 1376
QY 995 SSYEQCATRQPHVLDIEDFGMOGEGLGYSWMSISPRVDRVKNKNVPRRFRFGQGSVPREF 1054

```
Db 1377 SVSEDCS----- 1383
QY 1055 TGQIIPSTHELPGMGLSGSSSAVOEHQDDT-QHNQOENMKASHLOKTFDLNLSSEEC 1113
Db 1384 -----GLSSQSDILTTQORDTMQHN-----LIKLOQEMAELEAVLEQH 1421
QY 1114 LTRQSTKQNTDGCCLPRDRTAEDVVDPLNNSSLQNLVSNSSNKEQTAVEYKETNAT 1173
Db 1422 GSQPSNSVPSI-----ISDSSALEDLRNPEQSTSEKAVLTSQKSSEYP 1464
QY 1174 ILREMKGTADGKKPTSQWDSLRLKDVCEGQRGRNKNMDSIDYEAIRRASISEISEAI 1233
Db 1465 ISQNEGLSAD-KFEVSADSTSKNKE--PGVERSPSPKCPSLD----DRWYMHSCSGSL 1517
QY 1234 KERGM-NNMLAVRIKDFLERIVKDHGGIDL 1262
Db 1518 QNRNYPQOEELIKVVDVEEQOLEESGPHDL 1547
```

Search completed: July 5, 2001, 12:43:30
Job time: 73 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:18 ; Search time 16.84 Seconds
(without alignments)
3517.086 Million cell updates/sec

Title: PCT-US01-13059-2

Perfect score: 9089

Sequence: 1 MQSIMSSAVNATEFQND.....PRPLMARLHPASKLKNKT 1729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	246.5	2.7	3256	1 KI67_HUMAN	P46013 homo sapien
2	225.5	2.5	2375	1 ATRX_HUMAN	P46100 homo sapien
3	225	2.5	2468	1 MAPB_HUMAN	P46821 homo sapien
4	211.5	2.3	2842	1 APC_RAT	P70478 rattus norv
5	210.5	2.3	1805	1 NEST_RAT	P21263 rattus norv
6	208.5	2.3	3924	1 ANK2_HUMAN	Q01484 homo sapien
7	207	2.3	1723	1 AIM1_HUMAN	Q9V4K1 homo sapien
8	204.5	2.2	2717	1 ZEP1_HUMAN	P15822 homo sapien
9	200.5	2.2	3130	1 DPOZ_HUMAN	O60673 homo sapien
10	192.5	2.1	1781	1 AKAC_HUMAN	Q02952 homo sapien
11	192.5	2.1	1812	1 BRC1_MOUSE	P48754 mus musculus
12	192	2.1	2442	1 CBP_HUMAN	Q92793 homo sapien
13	191.5	2.1	2349	1 TPR_HUMAN	P12270 homo sapien
14	190	2.1	1284	1 ATL1_COWPX	P16602 cowpox viru
15	190	2.1	2464	1 MAPB_MOUSE	P14873 mus musculus
16	189	2.1	2453	1 NCR1_MOUSE	Q60974 mus musculus
17	188.5	2.1	1435	1 EBAL_PLAFC	P19214 plasmodium
18	188.5	2.1	3396	1 PGCV_HUMAN	P13611 homo sapien
19	186.5	2.1	1132	1 YKK5_YEAST	P34250 saccharomyc
20	186.5	2.1	1833	1 ZEP2_HUMAN	P31629 homo sapien
21	186	2.0	1531	1 NTF5_HUMAN	O94916 homo sapien
22	184.5	2.0	1875	1 MLP1_YEAST	Q02455 saccharomyc
23	183	2.0	2441	1 CBP_MOUSE	P45481 mus musculus
24	182.5	2.0	1861	1 MAP2_RAT	P15146 rattus norv
25	182	2.0	2688	1 ZEP1_MOUSE	Q03172 mus musculus
26	181	2.0	1658	1 YME7_YEAST	Q03661 saccharomyc
27	178	2.0	2230	1 GOG4_HUMAN	Q13439 homo sapien
28	178	2.0	2476	1 ATRX_MOUSE	Q61687 mus musculus
29	177.5	2.0	1972	1 P531_HUMAN	Q12888 homo sapien
30	177.5	2.0	2649	1 BPA1_HUMAN	Q03001 homo sapien
31	176.5	1.9	3358	1 PGCV_MOUSE	Q62059 mus musculus
32	175.5	1.9	2845	1 APC_MOUSE	Q61315 mus musculus
33	175	1.9	1210	1 AFX_HUMAN	P51825 homo sapien

34	174.5	1.9	1902	1 SMF1_HUMAN	O14497 homo sapien
35	174	1.9	728	1 TRDN_HUMAN	Q13061 homo sapien
36	173.5	1.9	839	1 GLT5_WHEAT	P10388 triticum ae
37	173.5	1.9	3210	1 CENF_HUMAN	P49454 homo sapien
38	173	1.9	2116	1 MYS2_DICDI	P08799 dictyostell
39	173	1.9	2414	1 P300_HUMAN	Q09472 homo sapien
40	173	1.9	3866	1 HRX_MOUSE	P55200 mus musculu
41	172.5	1.9	2843	1 APC_HUMAN	P25054 homo sapien
42	172.5	1.9	3828	1 TRX_DROVI	Q24742 drosophila
43	172	1.9	1863	1 BRC1_HUMAN	P38398 homo sapien
44	171	1.9	838	1 GLT4_WHEAT	P08489 triticum ae
45	171	1.9	2871	1 DESP_HUMAN	P15924 homo sapien

ALIGNMENTS

RESULT 1
KI67_HUMAN
ID KI67_HUMAN STANDARD; PRT; 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANTIGEN KI-67.
GN MKI67.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; PubMed=8227122;
RA Schlueter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
RN [2]

SEQUENCE OF 1-31 FROM N.A.
RP Gerdes J.;
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
CC PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC PHASE IN THE PERINUCLEOLAR REGION, IN THE LATER PHASES IT IS ALSO
CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY
CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL
CC CHROMOSOMES.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION OF THIS ANTIGEN OCCURS
CC PREFERENTIALLY DURING LATE G1, S, G2, AND M PHASES OF THE CELL
CC CYCLE, WHILE IN CELLS IN G0 PHASE THE ANTIGEN CANNOT BE DETECTED.
CC -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X65550; CAA46519.1; -
DR EMBL: X65551; CAA46520.1; -
DR EMBL: X94762; CAA64388.1; -
DR MIM: 176741; -
DR InterPro: IPR000253; -
DR Pfam: PF00498; FHA; 1.
DR PROSITE: P550006; FHA_DOMAIN; 1.
KW Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;

Db 3038 RCKSPVPVIMKRSR-----TSAKRIEPAEELNSNDM-KTNKEHKLODSVPENKGISLR 3092

QY 1532 LKNISRLRTEHQVEL-PDSHRLLDGMDKRPDDPSYLLAIWTPGETA----- 1579

Db 3093 SRRODKTEAQOITEFVLAIEINRNEKKPKMKTSP-MDIQNPDGDKGKPIPRDKVTE 3151

QY 1580 -----NSAQP-EQKCGKASGKCMCFD-----ETCSECNLSLRANSQTVR 1618

Db 3152 NKRCLRSARQNESSQPKVAESGGKSAKVLQMNQKGEAGNSDSMLRSRRTKSOPAA 3211

QY 1619 GTL 1621

Db 3212 STL 3214

RESULT 2

ATTRX_HUMAN

ID ATTRX_HUMAN STANDARD; PRT; 2375 AA.

AC P46100; P51068; Q15886;

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE TRANSCRIPTIONAL REGULATOR ATRX (X-LINKED HELICASE II) (X-LINKED

DE NUCLEAR PROTEIN) (XNP).

CN ATRX OR RAD54L OR XR2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-97123494; PubMed-8968741;

RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.,

RA Gibbons R.J.;

RT "ATR-X encodes a novel member of the SNF2 family of proteins: mutations

RT point to a common mechanism underlying the ATR-X syndrome.";

RL Hum. Mol. Genet. 5:1899-1907(1996).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE-97386582; PubMed-9244431;

RA Villard L., Lossi A.M., Cardoso C., Proud V., Chiaroni P.,

RA Collea L., Schwartz C., Fontes M.;

RT "Determination of the genomic structure of the XNP/ATR-X gene encoding

RT a potential zinc finger helicase.";

RL Genomics 43:149-155(1997).

RN [3]

RP SEQUENCE OF 743-2375 FROM N.A.

RX MEDLINE-95179111; PubMed-7874112;

RA Stayton C.L., Dabovic B., Gullisano M., Gecz J., Broccoli V.,

RA Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,

RA Bianchi M.E., Consalez G.G.;

RT "Cloning and characterization of a new human Xq13 gene, encoding a

RT putative helicase.";

RL Hum. Mol. Genet. 3:1957-1964(1994).

RN [4]

RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.

RX MEDLINE-94214473; PubMed-8162050;

RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.,

RA Millasseau P., Khrestchatsky M., Fontes M.;

RT "Cloning and expression of the murine homologue of a putative human

RT X-linked nuclear protein gene closely linked to PK1 in Xq13.3.";

RL Hum. Mol. Genet. 3:39-44(1994).

RN [5]

RP SEQUENCE OF 2284-2375 FROM N.A., AND VARIANTS ATR-X.

RX MEDLINE-95211835; PubMed-7697714;

RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;

RT "Mutations in a putative global transcriptional regulator cause X-

RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";

RL Cell 80:837-845(1995).

RN [6]

RP VARIANTS ATR-X.

RX MEDLINE-97467722; PubMed-9326931;

RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,

RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,

RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,

RA Higgs D.R.;

RT "Mutations in transcriptional regulator ATRX establish the functional

RT significance of a PHD-like domain.";

RL Nat. Genet. 17:146-148(1997).

RN [7]

RP VARIANT JM GLN-2014.

RX MEDLINE-96224392; PubMed-8630485;

RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Weber P.,

RA Munnich A., Lyonnet S.;

RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";

RL Nat. Genet. 12:359-360(1996).

CC -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES

CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN

CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-

CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X

CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE

CC PSYCHOMOTOR RETARDATION, CHARACTERISTIC FACIAL FEATURES, GENITAL

CC ABNORMALITIES, AND ALPHA-THALASSEMIA.

CC -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERGER-MARSIDI SYNDROME

CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY

CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,

CC MICROGENTALISM AND EARLY DEATH.

CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.

CC -!- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch)

CC -----

DR EMBL; U72938; AAB49971.1; -

DR EMBL; U72935; AAB40699.1; -

DR EMBL; U72904; AAB40699.1; JOINED.

DR EMBL; U72907; AAB40699.1; JOINED.

DR EMBL; U72908; AAB40699.1; JOINED.

DR EMBL; U72909; AAB40699.1; JOINED.

DR EMBL; U72910; AAB40699.1; JOINED.

DR EMBL; U72911; AAB40699.1; JOINED.

DR EMBL; U72912; AAB40699.1; JOINED.

DR EMBL; U72913; AAB40699.1; JOINED.

DR EMBL; U72914; AAB40699.1; JOINED.

DR EMBL; U72915; AAB40699.1; JOINED.

DR EMBL; U72916; AAB40699.1; JOINED.

DR EMBL; U72918; AAB40699.1; JOINED.

DR EMBL; U72919; AAB40699.1; JOINED.

DR EMBL; U72920; AAB40699.1; JOINED.

DR EMBL; U72921; AAB40699.1; JOINED.

DR EMBL; U72922; AAB40699.1; JOINED.

DR EMBL; U72923; AAB40699.1; JOINED.

DR EMBL; U72924; AAB40699.1; JOINED.

DR EMBL; U72925; AAB40699.1; JOINED.

DR EMBL; U72926; AAB40699.1; JOINED.

DR EMBL; U72927; AAB40699.1; JOINED.

DR EMBL; U72928; AAB40699.1; JOINED.

DR EMBL; U72929; AAB40699.1; JOINED.

DR EMBL; U72930; AAB40699.1; JOINED.

DR EMBL; U72931; AAB40699.1; JOINED.

DR EMBL; U72932; AAB40699.1; JOINED.

DR EMBL; U72933; AAB40699.1; JOINED.

DR EMBL; U72934; AAB40699.1; JOINED.

DR EMBL; U72936; AAB49969.1; -

DR EMBL; U72935; AAB40700.1; -

DR EMBL; U72908; AAB40700.1; JOINED.

DR EMBL; U72909; AAB40700.1; JOINED.

DR EMBL; U72910; AAB40700.1; JOINED.

DR EMBL; U72911; AAB40700.1; JOINED.

DR EMBL; U72912; AAB40700.1; JOINED.
 DR EMBL; U72913; AAB40700.1; JOINED.
 DR EMBL; U72914; AAB40700.1; JOINED.
 DR EMBL; U72915; AAB40700.1; JOINED.
 DR EMBL; U72916; AAB40700.1; JOINED.
 DR EMBL; U72917; AAB40700.1; JOINED.
 DR EMBL; U72918; AAB40700.1; JOINED.
 DR EMBL; U72920; AAB40700.1; JOINED.
 DR EMBL; U72921; AAB40700.1; JOINED.
 DR EMBL; U72922; AAB40700.1; JOINED.
 DR EMBL; U72923; AAB40700.1; JOINED.
 DR EMBL; U72924; AAB40700.1; JOINED.
 DR EMBL; U72925; AAB40700.1; JOINED.
 DR EMBL; U72926; AAB40700.1; JOINED.
 DR EMBL; U72927; AAB40700.1; JOINED.
 DR EMBL; U72928; AAB40700.1; JOINED.
 DR EMBL; U72929; AAB40700.1; JOINED.
 DR EMBL; U72930; AAB40700.1; JOINED.
 DR EMBL; U72931; AAB40700.1; JOINED.
 DR EMBL; U72932; AAB40700.1; JOINED.
 DR EMBL; U72933; AAB40700.1; JOINED.
 DR EMBL; U72934; AAB40700.1; JOINED.
 DR EMBL; X83753; CAAS8711.1; -.
 DR EMBL; U97103; AAC51657.1; -.
 DR EMBL; AF000157; AAC51657.1; JOINED.
 DR EMBL; AF000158; AAC51657.1; JOINED.
 DR EMBL; AF000159; AAC51657.1; JOINED.
 DR EMBL; AF000160; AAC51657.1; JOINED.
 DR EMBL; U97080; AAC51657.1; JOINED.
 DR EMBL; U97081; AAC51657.1; JOINED.
 DR EMBL; U97082; AAC51657.1; JOINED.
 DR EMBL; U97083; AAC51657.1; JOINED.
 DR EMBL; U97084; AAC51657.1; JOINED.
 DR EMBL; U97085; AAC51657.1; JOINED.
 DR EMBL; U97086; AAC51657.1; JOINED.
 DR EMBL; U97087; AAC51657.1; JOINED.
 DR EMBL; U97088; AAC51657.1; JOINED.
 DR EMBL; U97089; AAC51657.1; JOINED.
 DR EMBL; U97090; AAC51657.1; JOINED.
 DR EMBL; U97091; AAC51657.1; JOINED.
 DR EMBL; U97092; AAC51657.1; JOINED.
 DR EMBL; U97093; AAC51657.1; JOINED.
 DR EMBL; U97094; AAC51657.1; JOINED.
 DR EMBL; U97095; AAC51657.1; JOINED.
 DR EMBL; U97096; AAC51657.1; JOINED.
 DR EMBL; U97097; AAC51657.1; JOINED.
 DR EMBL; U97098; AAC51657.1; JOINED.
 DR EMBL; U97099; AAC51657.1; JOINED.
 DR EMBL; U97100; AAC51657.1; JOINED.
 DR EMBL; U97101; AAC51657.1; JOINED.
 DR EMBL; U97102; AAC51657.1; JOINED.
 DR EMBL; U98200; AAC50069.1; -.
 DR EMBL; L34363; AAA20872.1; ALT_SEQ.
 DR MIM; 300032; -.
 DR MIM; 301040; -.
 DR MIM; 309590; -.
 DR InterPro; IPR000330; -.
 DR InterPro; IPR001650; -.
 DR Pfam; PF00176; SNE2_N; 1.
 DR Pfam; PF00271; helicase_C; 1.
 KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
 KW Disease mutation; Alternative splicing.
 FT DOMAIN 103 151
 FT NP_BIND 1477 1484
 FT SITE 1602 1605
 FT DECH_BOX
 FT DOMAIN 628 633
 FT POLY-SER.
 FT DOMAIN 1034 1039
 FT POLY-LYS.
 FT DOMAIN 1049 1052
 FT POLY-SER.
 FT DOMAIN 1085 1089
 FT POLY-ASP.
 FT DOMAIN 1142 1149
 FT POLY-GLU.
 FT DOMAIN 1326 1349

Query Match 2.5%; Score 225.5; DB 1; Length 2375;

Best Local Similarity 18.8%; Pred. No. 0.00071;
 Matches 254; Conservative 187; Mismatches 510; Indels 401; Gaps 55;
 QY 27 LEPLNKTPOOKPSKRKRKMPK-VVVEGK---PKRPRKPAELPKVVV---EGKPRKRP 79
 DB 292 LEEDLNSEFRAMDAVNKEKNTKEHKVIDAKPETAKEGKPCALEKKDKISKSEAKLSRKQ 351
 QY 80 -----RKAATQEKVKSKETGSA-KKKNLKEKATKPPANVGD-----MSNKSPEVT 123
 DB 352 VDSEHMHQNVPTSEQRNKTSGGHEKKSDRKEEPOYEPANTSEDLMDIVSPSSVPEDI 411
 QY 124 LKSCRKALNF--DLENPGDARQGDSESEIVQNSGANSFSEIRDAIGTNGSFIDSVSQI 181
 DB 412 FENLETAMEVQSSVDHOGDSSG-TEQEVSSVKLAISS--KDNRGG----IKSKTTA 463
 QY 182 DKTNGLGAMNOPLEVSMGNOPDKLSTGAKLARDOOP-----DLTRNOOCOPPVATQNTQF 237
 DB 464 KVTIKELYVKLTP--VSLPNSPIKAGDCQEVQDRKDYKSCGLNPKLEKCGLGQENSDNEH 521
 QY 238 PMENQQAQLQMKNLIGLIFPGNQOP-----RWTIRNQOPC-LAMGNOQPMYLIQTPR 288
 DB 522 LVENEVSLLLEESDLRRSPRVKTTPLRRPTETNPVTSNDEECNETVKEKOKLSVPVRKK 581
 QY 289 PALVSGNQQLGGPOGKRRP-----
 DB 582 DKRNSDSAIDNPKNLPKSKQSETVDQNSDSEMLAILKGVSRMSSSSSDTDINEIH 641
 QY 310 LNHQTCPLPAGNOLYGSTDMHQLVMSTGGQOHLINQKQPGSLIRGQPCVPLIDQOPA 369
 DB 642 TNHKTLYDLTKTAGKDDKGRKRKSSSGSDFDTKKGSAKSSIIKKK-----ROTQS 695
 QY 370 TPKGFTHLNQMVATSMSSPGLRPHSQSOVPTTYLHVESVSRLNGTTGTCQRRAPAYDS 429
 DB 696 ESSNYDSELEKEIKSMKIGAAATTTKRIPNT-----KDFDS 732
 QY 430 LQODIH-----QGNKYI-LSHEISNGCGCKKALPONSS-----LPTPIM-----A 468
 DB 733 SEDEKHSKGMNQGHKNLTKSQEGSDDAERKQERETFSAGETVDKDTTMELRDLRP 792
 QY 469 KLEEARGSKRYHPRAMGOTEKHDLNLAQAQSQDQVHERHNSSTCEVYLDAAKTKIKQV 528
 DB 793 KKOQASASTDGVKLSKEQSFTSLEVKYVAETKEKSKH-----LKTVKTKCKV 840
 QY 529 QENLHGMPPPEVI-----EIEDDPTDGARKG---KNTASISKASK------GNSSPVK 572
 DB 841 QDGLSDIAEFLKQDSDETSEDDKQSKGKTEKKKPSDFKKVKIWEQYESSSDGTE 900
 QY 573 KTAEEK-CTVPK--TPAKKGRA-GRKKSVPPPAHASEIQLWQTPPKTPLS-----RSKP 624
 DB 901 KLPEREIECHFPKGIKQIKNGTTDGEKKKIRDKTSK-----RKDELSDYAERSTG 952
 QY 625 KKGKRSIQDSGARGPSGELLQODSAETIYRMQNLVLDGKEREQEQONAMVLYKGDGAL 684
 DB 953 KGDSCDSEDKSKNGAYR-----EKKRCK-----L 979
 QY 685 VPYESKRKRPRKVDIDDETRINWLMKGDEKGEDEKDKKKEKWEERERRVFRGRAD 744
 DB 980 LGRSSRRKQ-----DCSSSDTEKYSM-----KEDGCSNDDKRLKRIELRERNLSKRN 1028
 QY 745 SFIARMLVQDRRFRFPKSGVSDSVIGVFLTQNVSDHLSSAPMSLAARPPPKLSSSRE 804
 DB 1029 -----TKBIQSGSSS-----DAEESSE 1046
 QY 805 DER-----NVRSVVVEDEGCILNLEIPSMQEVQHPSDMEVSGVDSGSKQELRDC 856
 DB 1047 DNKKKORTSSKKKAVIVKEKK-----RNSLRTSTRKKQADITSSSSSDIEDD 1094
 QY 857 SNSGIERNFLEKSIQNLLEE-VLSSQDSFDPAIFQSCGRVSGSCSKS-----DA 906
 DB 1095 QNSIGEGSSDEQKIKPVTENLVLSHTGF-----CQSSGDEALSKSPVTVDDDDDD 1147
 QY 907 EFFTTRCETKTVSGTSQSQTGSPNLSDLEICLOGNERPHLYEGSGDVQ-----KQETNVA 962

Db 1148 NDPENRIAKKML-----LEEKANLSSD-----EDGSSDDPEEGKRTGKQN 1190

QY 963 QKKPDKMTNKKVSCVCPQPRNDNWTTPSSSYEQCATROPHVLDIEDFGMOGEGLY 1022

Db 1191 EENPGEAKNOVS-----ESDSSESKKPRYRHLR--HKLTVSD-GESE--- 1237

QY 1023 SWMSISPRVDRVKNVPRFRGCGSVPREPTGOIIPSTHELPGMGLSGSSSAVOHQ 1082

Db 1238 -----EKKTKPEHKEVGRNRK-----VSESESDSFQESGVSEVSESE 1280

QY 1083 DDTOHNOOD-----EMNKASHLQRTFLDLLNSSECLTROSSTKQNTDGLCPDRDAE 1136

Db 1281 DQRPRTSAAKAELEENQSVKQ-----KRR-----RIKVQE 1315

QY 1137 DVDPPLSNSSQLNTLVSNSSNKQTAVEYKETNATILREMKGTLAGCKPTSDWSLR 1196

Db 1316 D-----SSSEN---KSNSEEEEEEEEEEEEEEDENDDSDSPKGRKKIR 1364

QY 1197 KDVEGNEGRQERNKNMDSIDYEATRRASISE 1228

Db 1365 KILKDKLRT-----TONALKEEERKRRAIE 1392

RESULT 3

MAPB_HUMAN

AC P46821; MAPB_HUMAN STANDARD; PRT: 2468 AA.

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE MICROTUBULE-ASSOCIATED PROTEIN 1B [CONTAINS: MAP1 LIGHT CHAIN LC1].

GN MAP1B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95104835; PubMed=7806212;

RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;

RT "Cloning of human microtubule-associated protein 1B and the

RL identification of a related gene on chromosome 15.";

Genomics 22:273-280(1994).

CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.

CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES

CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST

CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS

CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN

CC STABILIZING MICROTUBULES.

CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE

CC WITH MAP1A AND MAP1B PROTEINS.

CC -!- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE

CC KKEE AND KKEI/V, REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER

CC REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES

CC BOTH IN VITRO AND IN VIVO.

CC -!- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED

CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH

CC BOTH MAP1A AND MAP1B.

CC -!- SIMILARITY: TO NEURAXIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: I06237; AAA18904.1; -.

DR MIM: 157129; -.

DR InterPro: IPR000102; -.

DR Pfam: PF00414; MAP1B_NEURAXIN; 10.

DR PROSITE: PS00230; MAP1B_NEURAXIN; 6.

KW Microtubules; Repeat; Phosphorylation.

FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.

FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY

FT KKEE AND KKEI/V REPEATS).

FT 12 X 17 AA TANDEM REPEATS.

FT DOMAIN 1869 2074 1.

FT REPEAT 1869 1885

FT REPEAT 1886 1902 2.

FT REPEAT 1903 1919 3.

FT REPEAT 1920 1936 4.

FT REPEAT 1937 1953 5.

FT REPEAT 1954 1970 6.

FT REPEAT 1971 1987 7.

FT REPEAT 1988 2004 8.

FT REPEAT 2005 2021 9.

FT REPEAT 2022 2038 10.

FT REPEAT 2039 2055 11.

FT REPEAT 2056 2072 12.

SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match 2.5%; Score 225; DB 1; Length 2468;

Best Local Similarity 18.5%; Pred. No. 0.00079;

Matches 336; Conservative 254; Mismatches 710; Indels 518; Gaps 81;

QY 175 LDSVSQIDKT---NGLGAMNOPLEVSNGNQDKLSTGAKLARDQDPLDLTRNQOCQRPV 230

Db 301 LDRVDSILLTHIGDDNLPGLNSMLQRTAELEESQSGSTTNSDMWNKLLISPDLVGVFLN 360

QY 231 ATQNTQFMENQOAWLQMKNLQIGFPFNGQNPRTIRNQOPL-AMGN----- 277

Db 361 VPENLKNPEPN---IKMRSIEEACFTLQYLNKLSMRPEPLFRSVGNITDPVILFKQK 416

QY 278 --QQPMYLIIGTPPA--LVSGNQQLGG-----POGKRPIFLNHQTC----- 316

Db 417 VGKLEMYVLNPKSKSEMOYFQWQGTGNKDKAEFLPNGQEVDPISPILTSVSSLLVWH 476

QY 317 PAGNQLGSPDTHQHVMSGTGQ-QHGLL-----IKN-----OQPSL---IRGQOCPVL 363

Db 477 PA-----NPAEKIIRVLPFGNSTYNILEGLEKHLKLDLFLKQPLATQKDLTGQVP-TPV 529

QY 364 IDQOPATPKGFTHLNMQ-----VATSMSSPGLRPHSQSQVP--TTLHVESVRL 412

Db 530 VKQ-----TKLQRADSRSLAPAKPLPSKSVKRESKEETPEVTKNHVE----- 575

QY 413 NGTTTCORSAPAYDSIQDDIHQGNKYLSHEISNGCKKALPQNSLSLPIPAKLEE 472

Db 576 -----KPKVESKEKVMVKDKPKVKE--TKPSVTEKEVPSKEP-PSPVKAEVAE 622

QY 473 ARGSKRQYHRAMQTEKHDLNLAQQAQSQ-----DVERHNSSTCVEYLDAAKTKIQ 526

Db 623 KOATDVKPAAKEKTVKKEKPKKEKKEKPKKEKKEKKEKKEKKEKKEKKEKKEKKEK 682

QY 527 VVOENLHGMPPVEIETDDPTDGARKGNTASISGASKGNSPVKKTAEKKEKIVPKTP 586

Db 683 EVKKEIKK-----BEKKEPKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKE 731

QY 587 AKKGGRAGKRSVPPPAHASEIQLWQTPP-----KTPLSRSKPKGKR-KSIQDSGKA 638

Db 732 IKKLPKDAKKSSTPLSEAKKPAALPKVPKPEESVKDSVAAGKPKKEGKIKVIRKEGKA 791

QY 639 RGPSELICQDSIAIIRYMONLYLGDREGEQONAVLYKGDGALVPYSEKRRPRPKV 698

Db 792 AEAVAAVGTGATTAATAVMAAAGIAAGPAKELEAERSLMSSPEDLTQKPEELKAE---EV 848

QY 699 DIDDFTTRINLLMGKDEKGEDEKDKKEK-----WVEERRVFRGRADSFIAHMLVQ 754

Db 849 DVTDKIKPQLELI-----EDEKLEKTEPEVYAVVQKEREVTKGPAESDEGITTTE 900

QY 755 GD-----RRFSPMKGSVYDVIGVFLTQNVSHLSSSAFMSLAARFPKPLSSREDER 807

Db 901 GEGECEQTPPEELEPVKEQGVDDI-----EKFEDEGAGF-EESSETGDVEE 944

QY 808 NVRSVVVEDDEG-----CILNLNIPSWQEK-----VOHFSDMEVSGVDSGSKEQ 852

Query Match 2.3%; Score 211.5; DB 1; Length 2842;
Best Local Similarity 18.4%; Pred. No. 0.0048;
Matches 319; Conservative 230; Mismatches 654; Indels 533; Gaps 83;

```
QY 64 AELPKVVE-----GKPRKRKAATOKVSKETGSAKKNLKESATKKPANV-----112
DB 883 AQIAKWEVEVSALHTSODDRSPASAEHLCHVAERTAA-----RSSASHPTNHNFAK 937
QY 113 GDMNKSPEVTL-----KSCRKALNFLENPCDARQDSESEIVQNSGANSFSEIRDA 166
DB 938 SESSNRCTSMYAKVEYKRSSNLSVTSDDGYGKRG-----QMKPSVESYEDDSG 990
QY 167 IGTNGSF-LDSVSDIKTNGL-----GAMNQPLEVSMGNQPKLSTG-----AKLARD 214
DB 991 KFCSYGOYPADLAHKIHSANHHDDNGEJDTPTINYSKYSDQLNSGROSPQNERWARP 1050
QY 215 QQ--PDLLTRNQOCFFVATOTQPMENQOAWLQMKNLQIGFPFGNQCPRTIRNQPC 272
DB 1051 KHVIEDEIKONEQRO--SRSONTFPVYSENT--DDKHLKFOOHFG-----QQEC 1096
QY 273 LAMGNOQPMYLLIGTPALVSGNOQLGGPQGNKRPIFLNHQTC-----LPAGNOLYG 324
DB 1097 VS-----PYRSGTNG-----SETNRMGSSHAVNQV--NQSLCQEDDYEDDKPTNYSERY 1145
QY 325 SPTDMHQ-----LWMTGGQOHLLIKNOQP-----GSLRGGQPCVPLIDQQA 369
DB 1146 SEEOHEEERPTNYSIKYNEEKHHV--DQPIDYSLKYATDISSSQKPSFS-FSKTFS 1200
QY 370 TPKGTHLQMWATSSPGLRPHSQSVPTTYLHVESVSRILNGTTCQSRAPAYDS 429
DB 1201 VOGTTEHNSPSEASAPSSNAKRQSQ-----LHPSSAQR--NGQTPGTACKVP--S 1250
QY 430 LQODIHQGNKYILS-----HEISNGN--GCKKALPONSLSPTIMAKLEEAR 474
DB 1251 INQETMQ--TYCVEDPTICFSCSSLSLSAEDELGCDDTQTEADSAANTLOIAEIKEND 1308
QY 475 GSKROYH-----RANGOTEKHLNLAAQIAQASODVERHNSCTVEYDAAK---KTKIQK 526
DB 1309 VTRSAQDPASDVAVSQSTRPSRLQASGLASARHKA---VEFSSGAKSPSKSGAQT 1365
QY 527 VVOENLHGMPPVEIEDDPTDGARKGKTATISKGASKNSPVYKKTAEKECI-----581
DB 1366 PKX-----PPHY--VOETPLVFSR-----CTSVS--SLDSFESRSTASSVQSEPCSGMVG 1413
QY 582 -----VPKTPAKGRAGRKSVPPPAHASEIQLMOPTPPKTPLRSKPKGKRKSIQDS 635
DB 1414 IVSPDLDPSPGQTWPPSRSKTTPPP-----PPQPVQTKREVPTKVPAEQR 1463
QY 636 GKARPSGELLQDSIAETIYRMQNLVLDGK-----EREQEQNAMVLYKGDGAL---VPYE 688
DB 1464 G---GPK-----QTAVSAAVQVQVLPDADTLHLHFATESTPDGFCSSLSALSDEPFI 1515
QY 689 SKRKRPR--PKVDIDDETTRINLLMGKDE-----KEGDEKDKKKKEKWEE-----735
DB 1516 QKDVELRIMPVOEND-----NGNETEPEQEESENQDKEVEKPDSEKDLDD 1564
QY 736 -----RRVER-----GRADSFTARMHLVQGDRR---758
DB 1565 SDDDDIEILEECIISAMPTKSRKAKKLAQATASKLPPPVARKPSQLPVYKLLPSQSRLOA 1624
QY 759 -----FSPWKGSVDSVIGVFTQNVSDHLSSAFMS-LAARFPKPLSSREDERNVRV 812
DB 1625 OKHVSFTPG-----DDVPRVYVEGTPIINFSTATSLDLTIESPPNELAAGD---GVRAS 1676
QY 813 VVEDPEGCTLNLNETPSWQEKVQHSDMEVSGVD-----SGSKEQ-----LRDCNSGIERF 864
DB 1677 V---OSGEFEKRTDITPEGRSTDEAQRGVSSIAIPDLDGSKAEEDGDLAECINSA----1729
QY 865 NFLEKSIONLEEVLSSQDSFDPATFQSG-----RV 896
DB 1730 --LPGRSHKPPRVKIMDQVOQAQSWTSSGNTKKNQIDTKKKKPTSPVKMPONTEYRTRV 1787
```

```
QY 897 GSCSCSKSDAEPTTTRCETKTVSGTSQSQVOTGSPNLSDELICLOGNE-----RPHL 946
DB 1788 RKNTDSKVNUNTEETFSDNK--DSKKQSLKNNPKLNDK--LPDNEDRVYRGGTFDPSPH 1843
QY 947 Y---EG-----SGDVQKQETTINVAOKKPDLEKTMNWKDSVCFQCPQRNDTNWOT 991
DB 1844 YAPIEGTPYCFRNDLSLSDFD-DDVDVLSREKAELRGKESKDS-----EAKVYCHT 1896
QY 992 TPSSSYE-----OCATRQPHVLDIEDFGMOGEGGLGYSWMSISPRVDRVKNKNVPRFRFQ 1046
DB 1897 EPPSSQOASARKAAQASTKH------VNRGPSKPLLEQEQTPFQ--SKDVPDR---1941
QY 1047 GGSVPREFTCQIIPSTPHELPGMGLSGSSSAVQEHODDTHNOODDMKNKASHLOKTFDL 1106
DB 1942 GAATDEKLFQFALENTP-----VCFSRNSSLSDSDVOEENNNNEETGPV-----1986
QY 1107 LNSSECLTRQSTQKITDGLCPDRDRTAEDVVDPILSNSSNQILVSNSSNKEQTAVE 1166
DB 1987 --RDAEPANAQOPGKPOASGYAPKSFHYEDTPVCFSRNSSLSSLSIDSEDD-----2036
QY 1167 YKETNATILRE-MKGTLDGKKPT-----SQWDSLRL-----KDVEGNEGRQ 1206
DB 2037 -----LLRECISAMPKKRRPSRLKGEQWQSPKVGSLAEDLTLDLKDQRPSEH 2089
QY 1207 ERKNKMSIDYEATPRASISEIAKERGMNMLAVRIKDFLERIVKDHGIDLEWL 1266
DB 2090 GLSPDS--ENFDKATQOEGANSIVSSLHQAASAAASQASDSDSILSKGSLGSPF 2148
QY 1267 ESPPOKADYLLSIRGLGLKSVCEVRLTLHLNLAFFVDTVNGVRIAVRMGWVPLQPLPESL 1326
DB 2149 HLTPDOEEKPTFSHG-----2164
QY 1327 QLHLLLELYVLESIOKFLWPRCLKDQRTLYELH-----YQITFGKVFC 1371
DB 2165 -----PRILKPEKSTLEAKKTESENKGIKGGKVKYKSLITSKI--2203
QY 1372 TKSRRNCN-ACPMRGECRHFASAYASARLALPAPERSTSATIPVPPSPFPVPAIPMIE 1430
DB 2204 ---RSNSELSSOMKOPLOTNMPISIRGMTIHIPGVNRNSSSTSPYSKKG-PP-----2252
QY 1431 LPLPLEKSLAG--AFPSNRENCEPIIEEPASQOECTEITESIEDAYYNEDP-----D 1482
DB 2253 LKTPASKSPSEGVATTSPRGTKPAVKSELSP-----ITROTSHISGSKNGPSRSGSRD 2306
QY 1483 EIPTIKLNIQEGMTLREHMERNMELQEGDMSKALVALHPTTTTPTP-CLKNISR 1537
DB 2307 STPS-----RPTQOPLSRPMQ-----SPGRNSISPGRNGISTPNKLSQLPR 2347

RESULT 5
NEST_RAT
ID NEST_RAT STANDARD; PRT; 1805 AA.
AC P21263;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE NESTIN.
GN NES.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90150286; PubMed=1689217;
RA Lendahl U., Zimmerman L.B., McKay R.D.G.;
RT "CNS stem cells express a new class of intermediate filament
protein."
RL Cell 60:585-595(1990).
CC -!- TISSUE SPECIFICITY: CNS STEM CELLS.
CC -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
```

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M34384; AAA41685.1; -
 DR PIR: A34736; A34736.
 DR InterPro: IPR001664; -
 DR Pfam: PF00038; Filament; 2.
 DR PROSITE: PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Heptad repeat pattern;
 KW Neurone.
 FT DOMAIN 1 7 HEAD.
 FT 8 314 ROD.
 FT 315 1805 TAIL.
 FT DOMAIN 8 43 COIL 1A.
 FT 44 55 LINKER 1.
 FT 56 151 COIL 1B.
 FT 152 174 LINKER 12.
 FT 175 193 COIL 2A.
 FT 194 196 LINKER 2.
 FT 197 314 COIL 2A.
 SQ SEQUENCE 1805 AA; 198744 MW; B40EE14717E0998D CRC64;

Query Match 2.38; Score 210.5; DB 1; Length 1805;
 Best Local Similarity 18.78; Pred. No. 0.0029;
 Matches 283; Conservative 216; Mismatches 564; Indels 447; Gaps 72;

Qy 60 PRKPAELPVVVEGKPKRKAATQEVKSK--ETGSAKKKLNKESATKRPANVGDSN 117
 Db 159 PRP-----PAPPHRIPGPAVEDELLARRLGEVWRGAVRDYQERVAAHMESSLGQARE 210
 Qy 118 KSEVTT--LKSCRKALNFDLENPGDARGDSEIIVQSSGANSFSEIRDAI--GGTNGSF 174
 Db 211 RLSQAVRGARECR-----LEVQQLQADRDLSLQERREALQRLGRW 251
 Qy 175 LDSVSDIKTN--GLGAMNQPLEVSMGNQPKLSTGAKLARDQDPDLTRNQOQFPVATQ 233
 Db 252 QDRLOATDKFQLAVEAEQERQGLQSQTAQLGGQQLAH-----LKMSLSEVATY 303
 Qy 234 NTQPFMENOQAWLQKMLGIFPGNQOPRMTIRNQOPLAMNQOPLYLIGTPPALVS 293
 Db 304 RTLLEAENSR--LQTPGRGSAISGLFDPKL-----KPNF--LGIPEDQYL--GSVLPALSP 354
 Qy 294 GNOQLGGPQGNKRPI--FLNHQTCIPAGNQLYSGPTDMHQLVMSTGGQHQGLLIKNOQPG 351
 Db 355 TSFSPPLNTLETPTVATFLKTQEFLOARTPTLAS--TPIPPISEAP-----CPN 402
 Qy 352 SLIRGQPCVPLIDQOATP--KGTHLNQWATSMSSPG-----LR 391
 Db 403 AEVRAQAEVPLSLOTAPEPLWATVPSSAILPELEPGKQKQGHFPDDLTLATNLN 462
 Qy 392 PHSOSQVPTVYLVHVESVRLNGTTCQSRAPAYDSLOQDIHQGNKYI-----LSHEI 446
 Db 463 PHH-----PT-----LEAKDGSESRVSSI--FOEDGOIWEVPEKADIEVKV 505
 Qy 447 SNGNGCKKALPQNSLPTPIIMAKLEEARG-----KROYHRAAGQTEKHDLNLAQOIAQSQ 502
 Db 506 ENSSAQK--TOESGLDT-----EEQDSQGPLOKETLKAIG--EPLMSLKIONYETA 554
 Qy 503 DVERHNST-----CVEYDRAKTKTKQVVQENL-----HGMV--PEVIEIDDPD 548
 Db 555 GKNCNSSTEGHGTGEPKEKEQIPLSLEKNVSEKTLLENGVPLSELLGKEDRTE 614
 Qy 549 -----GARK-----GKNTASISKASKGN-----SSPVKKT 574
 Db 615 DQELMSPKGLTKRFSGLSGSQEVVVRPSKEGNLESWTAFAKESOHPLGFGPAEDQMLERL 674

Qy 575 AEKEKICVPKTPAKKG-----RAGRKKSVPPPAH-----ASEIQLMQPTPPK 616
 Db 675 VEXEDQSPRSPEDDEACPLQKNEQPLGVEAEQILERLIEKESQESLSPED 734
 Qy 617 TPLSRSPKPKG---GRKSIDQSKARGPSELGELCQDSIAELIYRMQNLYLGDK--ERQOE 671
 Db 735 QEAGRSQKQENQPLGVEAEADQMLER-----LIEKESQESLQKSPENQRIKPLERENQ 789
 Qy 672 QNAMVLYKGDGALPYGSKRKPVKVDIDDETTRINLLMKGDKDEKGD--EEDKKKEK 730
 Db 790 KSLRYLEENQETTFPLESRNORPLRSLEVEEEOIRYKPLEKVSQDSGLSLAENQVPLR 849
 Qy 731 WVEEE-----RRVFRGRADSFIAHMLVQGDRRFPWKGVSVDVIGVFL-----TQNVS 780
 Db 850 YLEEDDCINKSLLEDKTHKSLGSLDRNGDSIIIP-----QESETQVSLRPEEEDQRIV 904
 Qy 781 DHLSSAFMSLAARPPKLSRSSREDERNV--RSVVVEDPECCILNLEIPSWQEKVQHP 838
 Db 905 NMLEKES-----QFSSRSSEEEQVMERSLEGENHE---SLSSVEKEDQME--S 949
 Qy 839 DMEVSGVDSG-----SKEOLRDCNSGIERFNFLEKSIONL-----EE 876
 Db 950 QLEKESQDSGKSLDESOETFGPLEKENAESLRLAGDQDEEQLEQETQTQTLRAVNEQ 1009
 Qy 877 EVLSSQDSFDP-----AIFQSCGRVSGCS--CKSDAEFPFTR--CETKTVSQGS-- 922
 Db 1010 MAVSPKEKVDPELPKPLGNDQEIARSLGKENQESLVSLKEKGIETVKSLETEIIEPLETA 1069
 Qy 923 -----QSVOTGSPNLSDLEICLOGNERPHLYE---GSGDVQKQETNVAKKPDLEKT 971
 Db 1070 EEDLRRKSIDTQEPMLSTEVARETPEPDEPPGSLGSDVDENRETTLSLEKESQESL 1129
 Qy 972 MNW-----KDSVCGFQPRNDTN-----WQTPPSSSYEQ----- 999
 Db 1130 GKWNVEYVEDSQOCLQVEEGLOEQHOESLREVQOELPSSGNOQRWEDVVEGKAVGQEA 1189
 Qy 1000 -CATRQPHVLDIEDFGMOGEG-----LCYSWM--SISPRVDRVKNKNVPR 1041
 Db 1190 PLATTGVGTEDKAEHLRGGGEEAAAGELLQDIVGEAWSLGSSPEKQORVPAEALDN 1249
 Qy 1042 RFFQGGSVPREFTGQIIPTPHELPGMLGSSSSAVQEHODD--TOHNOQDEMKNASHLQ 1100
 Db 1250 ---LEGGAL-----EVP---VAQSMPEVTERDEDAQAGEQDSIEVTIGLE 1289
 Qy 1101 --KTFDLNLSSECLTQSSSTKQNTDGCPLRDRTAEDVVDPLSNSSNQNLIVESNSS 1158
 Db 1290 AARTGLEL-----EQEVVGLDPRHFAEEAIPPSLGEESVK----- 1326
 Qy 1159 NKEQTAVEYKETNATILREMKGTIADGKKPTSQWDSLRKQVVEG--NEGQERKNKNMDSID 1217
 Db 1363 CQGHEE---SESMEGWEEEAAS-----LE--TSDHEGSDAQPOR--PPTEDEEG 1405
 Qy 1276 YLLSIRGLGLKSVQVRLTLHLNLAFPVDTVNGVRIARVMGWVLOPLPESLQHLLELYP 1335
 Db 1406 AQAALTAPGPKLLE-----PCSPILPILTAH--ELQP 1435
 Qy 1336 VLESIQKFLW 1345
 Db 1436 QAEGIQEAGW 1445
 RESULT 6
 ANK2_HUMAN STANDARD; PRT: 3924 AA.
 ID ANK2_HUMAN
 AC Q01484; Q01485;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ANKYRIN 2 (BRAIN ANKYRIN) (ANKYRIN B) (ANKYRIN, NONERYTHROID).

ANK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain stem;
RX MEDLINE=94075409; PubMed=8253844;
RA Chan W., Kordeli E., Bennett V.;
RT "440-kD ankyrinB: structure of the major developmentally regulated
RT domain and selective localization in unmyelinated axons.";
RL J. Cell Biol. 123:1463-1473(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Brain stem;
RX MEDLINE=91302466; PubMed=1830053;
RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT "Isolation and characterization of cDNAs encoding human brain
RT ankyrins reveal a family of alternatively spliced genes.";
RL J. Cell Biol. 114:241-253(1991).
RN [3]
RP REVISIONS.
RA Carpenter S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 463-495 FROM N.A.
RX MEDLINE=92009921; PubMed=1833308;
RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA Lux S.E., Ward D.C., Forget B.G.;
RT "Isolation and chromosomal localization of a novel nonerythroid
RT ankyrin gene.";
RL Genomics 10:858-866(1991).
CC -!- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-
CC SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN
CC BAND 4.2. TO NA-K ATPASE. TO THE LYMPHOCYTE MEMBRANE PROTEIN GP85,
CC AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND
CC DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO
CC THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;
CC THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC CELLS THROUGHOUT THE BRAIN.
CC -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC AND FUNCTION (POTENTIAL).
CC -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z26634; CAB42644.1; -;
DR EMBL; X56957; CA440278.1; -;
DR EMBL; X56958; CA440279.2; -;
DR EMBL; M37123; AAA62828.1; -;
DR PIR; S14533; S14533;
DR PIR; A39643; B39643.
DR PIR; B39643; B39643.
DR PIR; S14569; S14569.
DR HSP; Q00420; LMC.
DR MIM; 106410; -;
DR InterPro; IPR000488; -;
DR InterPro; IPR000906; -;
DR InterPro; IPR002110; -;
DR Pfam; PF00791; ZU5; 1.
DR Pfam; PF00023; ank; 22.

DR Pfam; PF00531; death; 1.
DR PROSITE; PS50088; ANK_REPEAT; 20.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW Phosphorylation; Multigene family.
FT REPEAT 63 92
FT REPEAT 96 125
FT REPEAT 129 158
FT REPEAT 162 191
FT REPEAT 193 220
FT REPEAT 232 261
FT REPEAT 265 294
FT REPEAT 298 327
FT REPEAT 331 360
FT REPEAT 364 393
FT REPEAT 397 426
FT REPEAT 430 459
FT REPEAT 463 492
FT REPEAT 496 525
FT REPEAT 529 558
FT REPEAT 562 591
FT REPEAT 595 624
FT REPEAT 628 657
FT REPEAT 661 690
FT REPEAT 694 723
FT REPEAT 727 756
FT REPEAT 760 789
FT REPEAT 793 822
FT REPEAT 825 854
FT REPEAT 857 886
FT REPEAT 889 918
FT REPEAT 921 950
FT REPEAT 953 982
FT REPEAT 985 1014
FT REPEAT 1017 1046
FT REPEAT 1049 1078
FT REPEAT 1081 1110
FT REPEAT 1113 1142
FT REPEAT 1145 1174
FT REPEAT 1177 1206
FT REPEAT 1209 1238
FT REPEAT 1241 1270
FT REPEAT 1273 1302
FT REPEAT 1305 1334
FT REPEAT 1337 1366
FT REPEAT 1369 1398
FT REPEAT 1401 1430
FT REPEAT 1433 1462
FT REPEAT 1465 1494
FT REPEAT 1497 1526
FT REPEAT 1529 1558
FT REPEAT 1561 1590
FT REPEAT 1593 1622
FT REPEAT 1625 1654
FT REPEAT 1657 1686
FT REPEAT 1689 1718
FT REPEAT 1721 1750
FT REPEAT 1753 1782
FT REPEAT 1785 1814
FT REPEAT 1817 1846
FT REPEAT 1849 1878
FT REPEAT 1881 1910
FT REPEAT 1913 1942
FT REPEAT 1945 1974
FT REPEAT 1977 2006
FT REPEAT 2009 2038
FT REPEAT 2041 2070
FT REPEAT 2073 2102
FT REPEAT 2105 2134
FT REPEAT 2137 2166
FT REPEAT 2169 2198
FT REPEAT 2201 2230
FT REPEAT 2233 2262
FT REPEAT 2265 2294
FT REPEAT 2297 2326
FT REPEAT 2329 2358
FT REPEAT 2361 2390
FT REPEAT 2393 2422
FT REPEAT 2425 2454
FT REPEAT 2457 2486
FT REPEAT 2489 2518
FT REPEAT 2521 2550
FT REPEAT 2553 2582
FT REPEAT 2585 2614
FT REPEAT 2617 2646
FT REPEAT 2649 2678
FT REPEAT 2681 2710
FT REPEAT 2713 2742
FT REPEAT 2745 2774
FT REPEAT 2777 2806
FT REPEAT 2809 2838
FT REPEAT 2841 2870
FT REPEAT 2873 2902
FT REPEAT 2905 2934
FT REPEAT 2937 2966
FT REPEAT 2969 2998
FT REPEAT 3001 3030
FT REPEAT 3033 3062
FT REPEAT 3065 3094
FT REPEAT 3097 3126
FT REPEAT 3129 3158
FT REPEAT 3161 3190
FT REPEAT 3193 3222
FT REPEAT 3225 3254
FT REPEAT 3257 3286
FT REPEAT 3289 3318
FT REPEAT 3321 3350
FT REPEAT 3353 3382
FT REPEAT 3385 3414
FT REPEAT 3417 3446
FT REPEAT 3449 3478
FT REPEAT 3481 3510
FT REPEAT 3513 3542
FT REPEAT 3545 3574
FT REPEAT 3577 3606
FT REPEAT 3609 3638
FT REPEAT 3641 3670
FT REPEAT 3673 3702
FT REPEAT 3705 3734
FT REPEAT 3737 3766
FT REPEAT 3769 3798
FT REPEAT 3801 3830
FT REPEAT 3833 3862
FT REPEAT 3865 3894
FT REPEAT 3897 3926
FT REPEAT 3929 3958
FT REPEAT 3961 3990
FT REPEAT 3993 4022
FT REPEAT 4025 4054
FT REPEAT 4057 4086
FT REPEAT 4089 4118
FT REPEAT 4121 4150
FT REPEAT 4153 4182
FT REPEAT 4185 4214
FT REPEAT 4217 4246
FT REPEAT 4249 4278
FT REPEAT 4281 4310
FT REPEAT 4313 4342
FT REPEAT 4345 4374
FT REPEAT 4377 4406
FT REPEAT 4409 4438
FT REPEAT 4441 4470
FT REPEAT 4473 4502
FT REPEAT 4505 4534
FT REPEAT 4537 4566
FT REPEAT 4569 4598
FT REPEAT 4601 4630
FT REPEAT 4633 4662
FT REPEAT 4665 4694
FT REPEAT 4697 4726
FT REPEAT 4729 4758
FT REPEAT 4761 4790
FT REPEAT 4793 4822
FT REPEAT 4825 4854
FT REPEAT 4857 4886
FT REPEAT 4889 4918
FT REPEAT 4921 4950
FT REPEAT 4953 4982
FT REPEAT 4985 5014
FT REPEAT 5017 5046
FT REPEAT 5049 5078
FT REPEAT 5081 5110
FT REPEAT 5113 5142
FT REPEAT 5145 5174
FT REPEAT 5177 5206
FT REPEAT 5209 5238
FT REPEAT 5241 5270
FT REPEAT 5273 5302
FT REPEAT 5305 5334
FT REPEAT 5337 5366
FT REPEAT 5369 5398
FT REPEAT 5401 5430
FT REPEAT 5433 5462
FT REPEAT 5465 5494
FT REPEAT 5497 5526
FT REPEAT 5529 5558
FT REPEAT 5561 5590
FT REPEAT 5593 5622
FT REPEAT 5625 5654
FT REPEAT 5657 5686
FT REPEAT 5689 5718
FT REPEAT 5721 5750
FT REPEAT 5753 5782
FT REPEAT 5785 5814
FT REPEAT 5817 5846
FT REPEAT 5849 5878
FT REPEAT 5881 5910
FT REPEAT 5913 5942
FT REPEAT 5945 5974
FT REPEAT 5977 6006
FT REPEAT 6009 6038
FT REPEAT 6041 6070
FT REPEAT 6073 6102
FT REPEAT 6105 6134
FT REPEAT 6137 6166
FT REPEAT 6169 6198
FT REPEAT 6201 6230
FT REPEAT 6233 6262
FT REPEAT 6265 6294
FT REPEAT 6297 6326
FT REPEAT 6329 6358
FT REPEAT 6361 6390
FT REPEAT 6393 6422
FT REPEAT 6425 6454
FT REPEAT 6457 6486
FT REPEAT 6489 6518
FT REPEAT 6521 6550
FT REPEAT 6553 6582
FT REPEAT 6585 6614
FT REPEAT 6617 6646
FT REPEAT 6649 6678
FT REPEAT 6681 6710
FT REPEAT 6713 6742
FT REPEAT 6745 6774
FT REPEAT 6777 6806
FT REPEAT 6809 6838
FT REPEAT 6841 6870
FT REPEAT 6873 6902
FT REPEAT 6905 6934
FT REPEAT 6937 6966
FT REPEAT 6969 6998
FT REPEAT 7001 7030
FT REPEAT 7033 7062
FT REPEAT 7065 7094
FT REPEAT 7097 7126
FT REPEAT 7129 7158
FT REPEAT 7161 7190
FT REPEAT 7193 7222
FT REPEAT 7225 7254
FT REPEAT 7257 7286
FT REPEAT 7289 7318
FT REPEAT 7321 7350
FT REPEAT 7353 7382
FT REPEAT 7385 7414
FT REPEAT 7417 7446
FT REPEAT 7449 7478
FT REPEAT 7481 7510
FT REPEAT 7513 7542
FT REPEAT 7545 7574
FT REPEAT 7577 7606
FT REPEAT 7609 7638
FT REPEAT 7641 7670
FT REPEAT 7673 7702
FT REPEAT 7705 7734
FT REPEAT 7737 7766
FT REPEAT 7769 7798
FT REPEAT 7801 7830
FT REPEAT 7833 7862
FT REPEAT 7865 7894
FT REPEAT 7897 7926
FT REPEAT 7929 7958
FT REPEAT 7961 7990
FT REPEAT 7993 8022
FT REPEAT 8025 8054
FT REPEAT 8057 8086
FT REPEAT 8089 8118
FT REPEAT 8121 8150
FT REPEAT 8153 8182
FT REPEAT 8185 8214
FT REPEAT 8217 8246
FT REPEAT 8249 8278
FT REPEAT 8281 8310
FT REPEAT 8313 8342
FT REPEAT 8345 8374
FT REPEAT 8377 8406
FT REPEAT 8409 8438
FT REPEAT 8441 8470
FT REPEAT 8473 8502
FT REPEAT 8505 8534
FT REPEAT 8537 8566
FT REPEAT 8569 8598
FT REPEAT 8601 8630
FT REPEAT 8633 8662
FT REPEAT 8665 8694
FT REPEAT 8697 8726
FT REPEAT 8729 8758
FT REPEAT 8761 8790
FT REPEAT 8793 8822
FT REPEAT 8825 8854
FT REPEAT 8857 8886
FT REPEAT 8889 8918
FT REPEAT 8921 8950
FT REPEAT 8953 8982
FT REPEAT 8985 9014
FT REPEAT 9017 9046
FT REPEAT 9049 9078
FT REPEAT 9081 9110
FT REPEAT 9113 9142
FT REPEAT 9145 9174
FT REPEAT 9177 9206
FT REPEAT 9209 9238
FT REPEAT 9241 9270
FT REPEAT 9273 9302
FT REPEAT 9305 9334
FT REPEAT 9337 9366
FT REPEAT 9369 9398
FT REPEAT 9401 9430
FT REPEAT 9433 9462
FT REPEAT 9465 9494
FT REPEAT 9497 9526
FT REPEAT 9529 9558
FT REPEAT 9561 9590
FT REPEAT 9593 9622
FT REPEAT 9625 9654
FT REPEAT 9657 9686
FT REPEAT 9689 9718
FT REPEAT 9721 9750
FT REPEAT 9753 9782
FT REPEAT 9785 9814
FT REPEAT 9817 9846
FT REPEAT 9849 9878
FT REPEAT 9881 9910
FT REPEAT 9913 9942
FT REPEAT 9945 9974
FT REPEAT 9977 10006
FT REPEAT 10009 10038
FT REPEAT 10041 10070
FT REPEAT 10073 10102
FT REPEAT 10105 10134
FT REPEAT 10137 10166
FT REPEAT 10169 10198
FT REPEAT 10201 10230
FT REPEAT 10233 10262
FT REPEAT 10265 10294
FT REPEAT 10297 10326
FT REPEAT 10329 10358
FT REPEAT 10361 10390
FT REPEAT 10393 10422
FT REPEAT 10425 10454
FT REPEAT 10457 10486
FT REPEAT 10489 10518
FT REPEAT 10521 10550
FT REPEAT 10553 10582
FT REPEAT 10585 10614
FT REPEAT 10617 10646
FT REPEAT 10649 10678
FT REPEAT 10681 10710
FT REPEAT 10713 10742
FT REPEAT 10745 10774
FT REPEAT 10777 10806
FT REPEAT 10809 10838
FT REPEAT 10841 10870
FT REPEAT 10873 10902
FT REPEAT 10905 10934
FT REPEAT 10937 10966
FT REPEAT 10969 10998
FT REPEAT 11001 11030
FT REPEAT 11033 11062
FT REPEAT 11065 11094
FT REPEAT 11097 11126
FT REPEAT 11129 11158
FT REPEAT 11161 11190
FT REPEAT 11193 11222
FT REPEAT 11225 11254
FT REPEAT 11257 11286
FT REPEAT 11289 11318
FT REPEAT 11321 11350
FT REPEAT 11353 11382
FT REPEAT 11385 11414
FT REPEAT 11417 11446
FT REPEAT 11449 11478
FT REPEAT 11481 11510
FT REPEAT 11513 11542
FT REPEAT 11545 11574
FT REPEAT 11577 11606
FT REPEAT 11609 11638
FT REPEAT 11641 11670
FT REPEAT 11673 11702
FT REPEAT 11705 11734
FT REPEAT 11737 11766
FT REPEAT 11769 11798
FT REPEAT 11801 11830
FT REPEAT 11833 11862
FT REPEAT 11865 11894
FT REPEAT 11897 11926
FT REPEAT 11929 11958
FT REPEAT 11961 11990
FT REPEAT 11993 12022
FT REPEAT 12025 12054
FT REPEAT 12057 12086
FT REPEAT 12089 12118
FT REPEAT 12121 12150
FT REPEAT 12153 12182
FT REPEAT 12185 12214
FT REPEAT 12217 12246
FT REPEAT 12249 12278
FT REPEAT 12281 12310
FT REPEAT 12313 12342
FT REPEAT 12345 12374
FT REPEAT 12377 12406
FT REPEAT 12409 12438
FT REPEAT 12441 12470
FT REPEAT 12473 12502
FT REPEAT 12505 12534
FT REPEAT 12537 12566
FT REPEAT 12569 12598
FT REPEAT 12601 12630
FT REPEAT 12633 12662
FT REPEAT 12665 12694
FT REPEAT 12697 12726
FT REPEAT 12729 12758
FT REPEAT 12761 12790
FT REPEAT 12793 12822
FT REPEAT 12825 12854
FT REPEAT 12857 12886
FT REPEAT 12889 12918
FT REPEAT 12921 12950
FT REPEAT 12953 12982
FT REPEAT 12985 13014
FT REPEAT 13017 13046
FT REPEAT 13049 13078
FT REPEAT 13081 13110
FT REPEAT 13113 13142
FT REPEAT 13145 13174
FT REPEAT 13177 13206
FT REPEAT 13209 13238
FT REPEAT 13241 13270
FT REPEAT 13243 13272
FT REPEAT 13245 13274
FT REPEAT 13247 13276
FT REPEAT 13249 13278
FT REPEAT 13251 13280
FT REPEAT 13253 13282
FT REPEAT 13255 13284
FT REPEAT 13257 13286
FT REPEAT 13259 13288
FT REPEAT 13261 13290
FT REPEAT 13263 13292
FT REPEAT 13265 13294
FT REPEAT 13267 13296
FT REPEAT 13269 13298
FT REPEAT 13271 13300
FT REPEAT 13273 13302
FT REPEAT 13275 13304
FT REPEAT 13277 13306
FT REPEAT 13279 13308
FT REPEAT 13281 13310
FT REPEAT 13283 13312
FT REPEAT 13285 13314
FT REPEAT 13287 13316
FT REPEAT 13289 13318
FT REPEAT 13291 13320
FT REPEAT 13293 13322
FT REPEAT 13295 13324
FT REPEAT 13297 13326
FT REPEAT 13299 13328
FT REPEAT 13301 13330
FT REPEAT 13303 13332
FT REPEAT 13305 13334
FT REPEAT 13307 13336
FT REPEAT 13309 13338
FT REPEAT 13311 13340
FT REPEAT 13313 13342
FT REPEAT 13315 13344
FT REPEAT 13317 13346
FT REPEAT 13319 13348
FT REPEAT 13321 13350
FT REPEAT 13323 13352
FT REPEAT 13325 13354
FT REPEAT 13327 13356
FT REPEAT 13329 13358
FT REPEAT 13331 13360
FT REPEAT 13333 13362
FT REPEAT 13335 13364
FT REPEAT 13337 13366
FT REPEAT 13339 13368
FT REPEAT 13341 13370
FT REPEAT 13343 13372
FT REPEAT 13345 13374
FT REPEAT 13347 13376
FT REPEAT 13349 13378
FT REPEAT 13351 13380
FT REPEAT 13353 13382
FT REPEAT 13355 13384
FT REPEAT 13357 13386
FT REPEAT 13359 13388
FT REPEAT 13361 13390
FT REPEAT 13363 13392
FT REPEAT 13365 13394
FT REPEAT 13367 13396
FT REPEAT 13369 13398
FT REPEAT 13371 13400
FT REPEAT 13373 13402
FT REPEAT 13375 13404
FT REPEAT 13377 13406
FT REPEAT 13379 13408
FT REPEAT 13381 13410
FT REPEAT 13383 13412
FT REPEAT 13385 13414
FT REPEAT 13387 13416
FT REPEAT 13389 13418
FT REPEAT 13391 13420
FT REPEAT 13393 13422
FT REPEAT 13395 13424
FT REPEAT 13397 13426
FT REPEAT 13399 13428
FT REPEAT 13401 13430
FT REPEAT 13403 13432
FT REPEAT 13405 13434
FT REPEAT 13407 13436
FT REPEAT 13409 13438
FT REPEAT 13411 13440
FT REPEAT 13413 13442
FT REPEAT 13415 13444
FT REPEAT 13417 13446
FT REPEAT 13419 13448
FT REPEAT 13421 13450
FT REPEAT 13423 13452
FT REPEAT 13425 13454
FT REPEAT 13427 13456
FT REPEAT 13429 13458
FT REPEAT 13431 13460
FT REPEAT 13433 13462
FT REPEAT 13435 13464
FT REPEAT 13437 13466
FT REPEAT 13439 13468
FT REPEAT 13441 13470
FT REPEAT 13443 13472
FT REPEAT 13445 13474
FT REPEAT 13447 13476
FT REPEAT 13449 13478
FT REPEAT 13451 13480
FT REPEAT 13453 13482
FT REPEAT 13455 13484
FT REPEAT 13457 13486
FT REPEAT 13459 13488
FT REPEAT 13461 13490
FT REPEAT 13463 13492
FT REPEAT 13465 13494
FT REPEAT 13467 13496
FT REPEAT 13469 13498
FT REPEAT 13471 13500
FT REPEAT 13473 13502
FT REPEAT 13475 13504
FT REPEAT 13477 13506
FT REPEAT 13479 13508
FT REPEAT 13481 13510
FT REPEAT 13483 13512
FT REPEAT 13485 13514
FT REPEAT 13487 13516
FT REPEAT 13489 13518
FT REPEAT 13491 13520
FT REPEAT 13493 13522
FT REPEAT 13495 13524
FT REPEAT 13497 13526
FT REPEAT 13499 13528
FT REPEAT 13501 13530
FT REPEAT 13503 13532
FT REPEAT 13505 13534
FT REPEAT 13507 13536
FT REPEAT 13509 13538
FT REPEAT 13511 13540
FT REPEAT 13513 13542
FT REPEAT 13515 13544
FT REPEAT 13517 13546
FT REPEAT 13519 13548
FT REPEAT 13521 13550
FT REPEAT 13523 13552
FT REPEAT 13525 13554
FT REPEAT 13527 13556
FT REPEAT 13529 13558
FT REPEAT 13531 13560
FT REPEAT 13533 13562
FT REPEAT 13535 13564
FT REPEAT 13537 13566
FT REPEAT 13539 13568
FT REPEAT 13541 13570
FT REPEAT 13543 13572
FT REPEAT 13545 13574
FT REPEAT 13547 13576
FT REPEAT 13549 13578
FT REPEAT 13551 13580
FT REPEAT 13553 13582
FT REPEAT 13555 13584
FT REPEAT 13557 13586
FT REPEAT 13559 13588
FT REPEAT 13561 13590
FT REPEAT 13563 13592
FT REPEAT 13565 13594
FT REPEAT 13567 13596
FT REPEAT 13569 13598
FT REPEAT 13571 13600
FT REPEAT 13573 13602
FT REPEAT 13575 13604
FT REPEAT 13577 13606
FT REPEAT 13579 13608
FT REPEAT 13581 13610
FT REPEAT 13583 13612
FT REPEAT 13585 13614
FT REPEAT 13587 13616
FT REPEAT 13589 13618
FT REPEAT 13591 13620
FT REPEAT 13593 13622
FT REPEAT 13595 13624
FT REPEAT 13597 13626
FT REPEAT 13599 13628
FT REPEAT 13601 13630
FT REPEAT 13603 13632
FT REPEAT 13605 13634
FT REPEAT 13607 13636
FT REPEAT 13609 13638
FT REPEAT 13611 13640
FT REPEAT 13613 13642
FT REPEAT 13615 13644
FT REPEAT 13617 13646
FT REPEAT 13619 13648
FT REPEAT 13621 13650
FT REPEAT 13623 13652
FT REPEAT 13625 13654
FT REPEAT 13627 13656
FT REPEAT 13629 13658
FT REPEAT 13631 13660
FT REPEAT 13633 13662
FT REPEAT 13635 13664
FT REPEAT 13637 13666
FT REPEAT 13639 13668
FT REPEAT 13641 13670
FT REPEAT 13643 13672
FT REPEAT 13645 13674
FT REPEAT 13647 13676
FT REPEAT 13649 13678
FT REPEAT 13651 13680
FT REPEAT 13653 13682
FT REPEAT 13655 13684
FT REPEAT 13657 13686
FT REPEAT 13659 13688
FT REPEAT 13661 13690
FT REPEAT 13663 13692
FT REPEAT 13665 13694
FT REPEAT 13667 13696
FT REPEAT 13669 13698
FT REPEAT 13671 13700
FT REPEAT 13673 13702
FT REPEAT 13675 13704
FT REPEAT 13677 13706
FT REPEAT 13679 13708
FT REPEAT 13681 13710
FT REPEAT 13683 13712
FT REPEAT 13685 13714
FT REPEAT 13687 13716
FT REPEAT 13689 13718
FT REPEAT 13691 13720
FT REPEAT 13693 13722
FT REPEAT 13695 13724
FT REPEAT 13697 13726
FT REPEAT 13699 13728
FT REPEAT 13701 13730
FT REPEAT 13703 13732
FT REPEAT 13705 13734
FT REPEAT 13707 13736
FT REPEAT 13709 13738
FT REPEAT 13711 13740
FT REPEAT 13713 13742
FT REPEAT 13715 13744
FT REPEAT 13717 13746
FT REPEAT 13719 13748
FT REPEAT 13721 13750
FT REPEAT 13723 13752
FT REPEAT 13725 13754
FT REPEAT 13727 13756
FT REPEAT 13729 13758
FT REPEAT 13731 13760
FT REPEAT 13733 13762
FT REPEAT 13735 13764
FT REPEAT 13737 13766
FT REPEAT 13739 13768
FT REPEAT 13741 13770
FT REPEAT 13743 13772
FT REPEAT 13745 13774
FT REPEAT 13747 13776
FT REPEAT 13749 13778
FT REPEAT 13751 13780
FT REPEAT 13753 13782
FT REPEAT 13755 13

Db 2336 AQESTATSDTKALPLPEASVYKTDGTGTSKPGQVIRSPQGLEALPLPSRDSSEVL SAVADDS 2395
 QY 291 LVSGNQQLGGPOGNRPFLFHNQTCPLAG-----NOLYGSPTDMHQLVMSGGQHQGLLI 345
 Db 2396 LA-----VSHKSLASPLVDNNSHKTPDLSL----- 2422
 QY 346 KNOQPGSLIRGOQPCVPLIDQOPATPKGFTHLNQVATSMSPGLRPHRSQOVPTTYLHV 405
 Db 2423 ---EPSPL--KESPCRODLESSPVEPK-----MKAGIFP-SHPPLPAVAKT 2463
 QY 406 ESV-----SRILNGTGTQCORAPAYDSLOD--IHOGNKYLSHEI--SNGNGCKKA 455
 Db 2464 ELLTEVASVRKLLRDPDGSAD-----DSLEQTSMESSGKSPSPDTPSSEVSVEV 2517
 QY 456 LPONSSLPPTIMAKLEA-----RGSKROYHRAMGOTEKHDL-----NLAQQ 497
 Db 2518 TPKTTDVSTPAVIVHECAEEDDSENGEKKRP-----TPEEMFKWVTKMKFDELEQE 2571
 QY 498 IAQSDV-----ERHNSST-----CVELDAAKTKTKQVQVQENLHGMPEVIEIEDDP 546
 Db 2572 AKQKDYKKEPKQEBESSSSDPDADCSVDVDEPKHTG-----SGEDSGVPLVTSSEK 2627
 QY 547 TDGARKGNTASISKGASKG-----NSPVPKTAKEKCIYVKP- 584
 Db 2628 SSSSESEPLAQLKKGADSGLLPEVIRVQPPSPPLPSMDSNSPVEVQFQP-----VWSK 2684
 QY 585 -----TPAKKGRAGRKKS-----VPPAHASEIQWLQWTPPKTPLRSRKP-----KGR 629
 Db 2685 YTFKMNEDQEEPKSEKESSEHSLADRAVSTEAEDRSYDKLNRDTPQFKICDHGC 2744
 QY 630 KSIDQSGKAR-----GPSGELLQCDISAE--IYRMONLYLGDKEREQONAMVLYK 680
 Db 2745 EAMSPSSARPVSSGLOSPG-----DDVDEQPVYIKESLALQGTHEKDTGEEL----- 2794
 QY 681 DGALVPYESKRRKRPKVDIIDE-----TTRIWNILMGKDGKEGD----- 721
 Db 2795 -----DVSRAESPQADCPSESFSSSLPHGLVSEKDELDEISATSIQKTEVTKT 2846
 QY 722 EEKDKKKKEWEEERRVFRGRADF--IARHLVQGDRESPKGSVVD----- 769
 Db 2847 DETFENLPKDCPSQDSITTTQDRFSDMDVPVSDLAENDIYDQITSPYENVPQSFPSS 2906
 QY 770 -----VIGVELTQNVSDHL-----SSSAFMSLAARF----- 795
 Db 2907 EESKTQTDANHTTFSHSEVSVTITSPEDVVVASSSGTIVLSKESFEGQDIKMSQL 2966
 QY 796 -----PPLKSSREDERNVRSVV-EDPEGCILNLEI--PSWQE-----KV 834
 Db 2967 ESTLWEMQSDSVSSSEFPTATTVVGEQISKVITKTDVSDSWSEIREDDAEFAEV 3026
 QY 835 QHPSDMEVSG--VDSGSK-----EQ-----LRDCNSNGI-- 861
 Db 3027 KE-EQKIFGLMVDQSGTTPDTPARTPEETPTSPSEQNPFLFQEGKLFEMTRSGAID 3085
 QY 862 -----ERENFL-----EKSIONLEEVLSSQSDFPAPFQSGRVCSCSKSDA--- 906
 Db 3086 MTKRSYADESPHFQIGQESREETLSEDDYKBGATGADPLPLETSAESLALSKEKTVDD 3145
 QY 907 -----BPTTRCTKTVSGTSQ-----VOTGS-PNLS- 933
 Db 3146 ADLLPDSVSEVEEIPASDAQLNSQMGISASTETPTKEAVSVGTGKDLPTVOTGDIPLSG 3205
 QY 934 -DEICLQGNERP-----HLVEGSGDVQKQETTNVAOKKPDLEKTMNKDVSVC 979
 Db 3206 VKQISCDSSEPAVOQLDFSTLRSVYSDRGD--DSPDSSPEQKSVKIEPTAPMENVP 3263
 QY 980 FGQPRN-----DTNWOTTPSSSYEQCATRQPHVLDIEDFGMGEGLGYSWMSISPRVD 1032
 Db 3264 FTESKSKIPVMTPTSTPAPSAEYSSVS-----EDF-----LSSVDEENKAD 3307
 QY 1033 RVKKN-----VPRFRFQ-----GGSVPREFT--GQIIPSTPHELPGMGLSGSSVAQE 1080

RESULT 7
 AIM1 HUMAN
 ID AIM1_HUMAN STANDARD; PRT; 1723 AA.
 AC Q914KI; O00296;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ABSENT IN MELANOMA 1 PROTEIN.
 GN AIM1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=97250519; PubMed=9096375;
 RA Ray M.E., Wistow G., Su Y.A., Meltzer P.S., Trent J.M.;
 RT "AIM1, a novel non-lens member of the betagamma-crystallin
 superfamily, is associated with the control of tumorigenicity in human
 malignant melanoma.";
 RL Proc Natl Acad Sci U S A. 94:3229-3234(1997).
 CC -I- FUNCTION: MAY FUNCTION AS SUPPRESSOR OF MALIGNANT MELANOMA. IT MAY
 CC EXERT ITS EFFECTS THROUGH INTERACTIONS WITH THE CYTOSKELETON.
 CC -I- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY. CONTAINS
 CC 6 TANDEM REPEATS OF A BETA/GAMMA-TYPE CRYSTALLIN DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its


```

RA Omichinski J.G., Clore G.M., Robien M., Sakaguchi K., Appella E.,
RA Gronenborn A.M.;
RA "high-resolution solution structure of the double Cys2His2 zinc
RA finger from the human enhancer binding protein MBP-1.";
RL Biochemistry 31:3907-3917(1992).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC 5'-GGGCACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC IN T-CELL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
CC ZINC-FINGER IN-BETWEEN.
CC -1- SIMILARITY: STRONG, TO HIVP2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51435; CAA35798.1; -.
DR PIR; A34203; A34203.
DR PDB; 3ZNF; 15-JAN-92.
DR PDB; 4ZNF; 15-JAN-92.
DR PDB; 1BBO; 31-OCT-93.
DR TRANSFAC; T00497; -.
DR MIM; 194540; -.
DR InterPro; IPR000822; -.
DR Pfam; PF00096; zf-C2H2; 5.
DR PRINTS; PR00048; ZINC_FINGER.
DR PROSITE; PS00028; ZINC_FINGER_C2H2.1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2.2; 4.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Repeat; 3D-structure
FT DOMAIN 406 456 ZINC FINGERS.
FT ZN_FING 406 428 C2H2-TYPE.
FT ZN_FING 434 456 C2H2-TYPE.
FT DOMAIN 803 806 POLY-SER.
FT ZN_FING 958 981 C2HC-TYPE (POTENTIAL).
FT DOMAIN 2087 2139 ZINC FINGERS.
FT ZN_FING 2087 2109 C2H2-TYPE.
FT ZN_FING 2115 2139 C2H2-TYPE.
FT STRAND 2088 2088
FT TURN 2090 2092
FT STRAND 2095 2095
FT HELIX 2099 2108
FT TURN 2109 2109
FT STRAND 2115 2116
FT STRAND 2123 2124
FT HELIX 2127 2135
SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;
Query Match 2.2%; Score 204.5; DB 1; Length 2717;
Best Local Similarity 16.9%; Pred. No. 0.01;
Matches 355; Conservative 297; Mismatches 706; Indels 741; Gaps 93;
QY 14 EATEQNDS-----SQDYLEFLNLTPQ-QFSKRRKFRMPKVYVEGPKRKRPRKPAELPKV 69
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 18 EAQKELNGAEVSKKEILQAGVKGTSLSKGVKR-----KIVAENHLKTKPSPL----- 67
QY 70 VVEGPKRKRPRKRAATQEVKSKETGSAKKKNLSEATKKPANVGDMSNKSPEVTLASCRK 129
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 68 -----RNPLOA-----KHKQNTSESPVLHVSASESHKKNQNYIPVKNGKQ 107
QY 130 ALNFDLENPGDARQGGSEISEIVONSGANSFSETRDAIGTNGSFLDSVSQIDKTNGLGA 189

```

[illegible]


```
FT CONFLICT 237 237 /FTid=VAR_008519.
FT CONFLICT 1156 1156 E -> Q (IN REF. 4 AND 5).
FT CONFLICT 3130 3130 Y -> C (IN REF. 4 AND 5).
SQ SEQUENCE 3130 AA; 352782 MW; 1C0700900F10BB14 CRC64;

Query Match 2.2%; Score 200.5; DB 1; Length 3130;
Best Local Similarity 18.3%; Pred. No. 0.02;
Matches 345; Conservative 229; Mismatches 647; Indels 663; Gaps 88;

QY 6 DSSAVNATEATEQNGSRQDVLFDLNTKTPQQRKPKRMPKVVVGGPKRKRPR----- 61
Db 511 DNLULLASLIPQDGTADENSNDPLNNSRTHS-----SVIATSKLSVKPISFHK 562
QY 62 -----KPAELPKVVVEGPKRKRKAATQEKVKSKE-----TGSAKKKNKLESAT-- 106
Db 563 DAATLEFSSSAKITFOCK-----HTSALSHLVNLKEDLIEDLSOTNKANTKGLDNTSFTS 617
QY 107 -----KRPANVGD-----SNKSPEVTLKSCRKALNFLEN--PGDARQGD-- 146
Db 618 TNESTYSMKYPGSLSSVHSNSHSHKSKKEILLPVSSCESSI-FDYEDIPSVTRQVPSR 676
QY 147 -----ESELV-----QNSCANSESEIRDAIGTNGSFLDSVQIDKTNGLGA 189
Db 677 KYTNIRKIEKDSPPFIHHRHPNENTLGNFS-----NFDLHNSKNKVS----- 720
QY 190 MNOPLVSMGNOPDKLSTGAKLARDQOPDLLTRNOQCOFFVATONTOPMENOQAWLQMK 249
Db 721 -----SEGNEKNSALSLF-----PSEFTENCCELLSSCGENRTVMVHLSNSTADESGL 769
QY 250 NQLIGFFPGNOQPRMTIRNOQPCILAMGNOQPMYLLIGTPRALVSGNOOLGGP----- 302
Db 770 NKL-KIRVEEFOEHT--EKPSLSQAAAHVMP-----PSVVLN-CLTRPQKLSPTY 819
QY 303 -----GNKRPIFLNHTCLPAGNOLYGPDMHQLVMGTGGQHQGLLKNQOPGSLIRGQ 358
Db 820 KLOPGNK-PSRLKLNKRLAGHQTSTKSS-----BTGSKDNFIQNN----- 861
QY 359 PCVLIDQOPATPKGFTHLQMWATSSPCLRPHSOSQVPTT-----YLHVESV 408
Db 862 PCNSNPEKDNLASDLTKTT-----RGAFENKTPDGFIDCHFGDGTLETEQ- 908
QY 409 SRLINGTTGQRSAPAY---DSLQDDIHGNNKYILSHSISNGNG-----CKKALPON 459
Db 909 SFGLYGNKYTLRAKKVNYETEDSESSFVTHNSKISLPHPMEIGESLDGTLKSRKRKMS 968
QY 460 SSLPTPIAK--LEEARGSKROYHRAMGOTEKH-----LNLAAQOIAOSOV- 504
Db 969 KKLBP-PVLIKIYIINRPRGRKNMLVK-LGKIDSKEKOVILTEEMELYKKLAPLKDQFWPK 1026
QY 505 -----ERNSSTCVEYLDAAKKT-KIQKVVOENLHGMPPVEIEIDDP 546
Db 1027 VPDSPATKPIYIPLTPKSHRRKS--KHSAKKTGKQOQRTNENI----- 1070
QY 547 TDGARKGNKTASTKASKNSPVKKT--AEKEKCIIVKTPAKKGAGRKKSVPVPPAHA 604
Db 1071 -----KRTLSFRKKRSHAILSPSPSYNAETEDCDLNTSDVMSKGLGLSERSTSPINS 1123
QY 605 SEIQWQOPTPKTPLSRKPKGKRKSIQDSKARGSPGELLCCODSTAELIYRMONLY-- 662
Db 1124 SPPRCWSPDPR-----AEEIMAAAEKAMLFKGPVNYKK 1158
QY 663 -----LGDKREOQONAWLYKGDGALV--PYBSKKRPRPKVIDDETTRINLLMGK 715
Db 1159 TVNSRICTSRAQO-----IKSKAKLANPSIVTKKNNK-----NOTNKLVD----- 1202
QY 716 DEKBGDEKKKKEKWEERVRFRGRADSFIAHMLHVQDGRFRSPKWSVDSVIGVFL 775
Db 1203 DGKKKPRAKOKTNEK-----TSRKHITLDEKIKSOGAEVKFVL---K 1244
QY 776 TONVSDHLSSAFMSLAAREFPKLSRSSREDERNVRSVVVEDPECCILNLNEIPSWQEKVQ 835
Db 1245 HONVSEFASGGSGSL-----LPFKQKDMPLMGSAVD 1275
```



```
QY 730 KWEERRVFRGRADSFIAHMLVQGRDRFSPWKGSVVDSVIGVFLTONVSDHLSAFA 789
Db 639 -----SALISSTE-- 646
QY 790 SLAARFPKLSRREDERNVRSVVEDEPGICILNLNIPISQEKV-----QHPS 838
Db 647 STASEQEMKGSVEEPK-----PEEPKRKVDTSVSWEALICVGSKKRRARRSS 697
QY 839 DMEVSGVDSGSKQLRDCNSGIERFNFLKSTONLEEEVLSSQSDSDPAIFQSCGRVGS 898
Db 698 DEEGGPKAMGDGHOKAD--BAG-----KDKETGTDGLAGSQEHPD-----GQ 738
QY 899 CSKSKSDAEFPPTTCETKTCTVTSQSQTGSPNLSDBICLQGNRPHLYSGSGDVQKQET 958
Db 739 GSSSPEQAGSPT---EGEGVSTWESFKRLVTPRKSKSKLEEKSEDSI--AGSG----- 787
QY 959 TNVAQKPDLEKTNWKNKDSVCFOPRNDTNW-----OTTPSSVEQCATROPHY 1007
Db 788 --VEHSTPDTE-----PGKEESVWSIKKPIPRKKRKPDPGKQEQA----- 825
QY 1008 LDIEDFCMQGGLGYSWM--SISP--RVDRVKNKNVPRFFRQGSVPR-----EFTGQII 1059
Db 826 -PVEDAGTGANEDSDVPVAVPVLSEYDAVEREKMEAQQAQKGAQPEQKAAATEVSKELS 884
QY 1060 PSTPHELFGMLGSGSSA--VOEHQDDTOHNOQDMNKASHLQKTFLLDLSNSEECLTRQ 1117
Db 885 ESQVHMAAAVADGTRAATIEERSPSN-----ISASVTEPLEQV 924
QY 1118 SSTQNTDGLCPDRDAED-----VVDPLNSSLQ--NILVSNSSNKKOTAVEYKTN 1171
Db 925 EAAALITEEVLEVEVIAESEPPTVTPELPENREARGDTVVSEAEITPEAVTAAE---T 980
QY 1172 ATILREKMGTLADGKPKTSOWDSL-----RKDVEGNEGROERKNKNMDS 1215
Db 981 APLGSEGTETASAEETEMKWSAVSOLTSDPTTEATPVQVEGVPDIEQERTQE 1040
QY 1216 IDYBAIRRASISEISEAIKER---GMNNMLAVRIKFLERIVKDHGGIDLEWLRESPPD 1271
Db 1041 V-----LQAAVKEESQSLPGTGGP-----EDVLQPVQRAEA-----ERPEE 1078
QY 1272 KADYLLSIRGLGKSEVCVRLTLNLAFVDTNVGRI--AVRMGWVPLPPLPESLOLHL 1330
Db 1079 QAE-----ASGLKKET-----DVVLKVDQAQEAKEPFTQGVVGQTTPESEF-- 1120
QY 1331 LELPVLESIQKFLWPLRCLKDORTLYELHYOLITFGKVFCTKSRPNACNACPMRGECH 1390
Db 1121 -KAPQVTESIES-----SELVT-----TCQAEFLAG----- 1145
QY 1391 ASAYASARLALPAPEERSLTSATIPVPPFPVPAIPMIELPLP-----LEKS 1438
Db 1146 ---VKSQEMVM---EQAIIPDSVETPTDSETDGTSTPVADFAPGTQKDEIVEIHEENE 1198
QY 1439 LAGAPSNRENCEPI-----IEEPASPG-----QECTETESDIEDAYVNEPDP-EIPTKL 1489
Db 1199 VASQTSGGTAEAAVPAQKRPAPSPFVQEEFK-BQSKMEDTLEHTDKREVSVEVTSI 1256

RESULT 11
BRCL_MOUSE STANDARD; PRT; 1812 AA.
AC P48754; Q60957; Q60983;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE BREAST CANCER TYPE 1 SUSCEPTIBILITY PROTEIN HOMOLOG.
GN BRCA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=96177659; PubMed=8634697;
RA Abel K.J., Xy J., Yin G.Y., Lyons R.H., Meisler M.H., Weber B.L.;
RT "Mouse Brca1: localization sequence analysis and identification of
RL evolutionarily conserved domains.";
RN Hum. Mol. Genet. 4:2265-2273(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=96177660; PubMed=8634698;
RA Sharan S.K., Wims M., Bradley A.;
RT "Murine Brca1: sequence and significance for human missense
RN mutations.";
RN Hum. Mol. Genet. 4:2275-2278(1995).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=96121367; PubMed=8575748;
RA Bennett L.M., Haugen-Strano A., Cochran C., Brownlee H.A.,
RA Fiedorek F.T. Jr., Wiseman R.W.;
RT "Isolation of the mouse homologue of BRCA1 and genetic mapping to
RN mouse chromosome 11.";
RN Genomics 29:576-581(1995).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ; TISSUE=Embryo;
RX MEDLINE=96067162; PubMed=7590247;
RA Lane T.F., Deng C., Elson A., Lyu M.S., Kozak C.A., Leder P.;
RT "Expression of Brca1 is associated with terminal differentiation of
RL ectodermally and mesodermally derived tissues in mice.";
RN Genes, Dev. 9:2712-2722(1995).
[5]
RP SEQUENCE OF 727-1111 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=96021028; PubMed=7550308;
RA Marquis S.T., Rajan J.V., Wynshaw-Boris A., Xu J., Yin G.Y.,
RA Abel K.J., Weber B.L., Chodosh L.A.;
RT "The developmental pattern of Brca1 expression implies a role in
RL differentiation of the breast and other tissues.";
RN Nat. Genet. 11:17-26(1995).
[6]
RP SEQUENCE OF 789-1250 FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=96163506; PubMed=8566965;
RA Schroeck E., Badger P., Larson D., Erdos M., Wynshaw-Boris A.,
RA Ried T., Brody L.;
RT "The murine homolog of the human breast and ovarian cancer
RL susceptibility gene brca1 maps to mouse chromosome 11D.";
RN Hum. Genet. 97:256-259(1996).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. INVOLVED IN
CC TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE (BY
CC SIMILARITY).
CC -!- SUBUNIT: CHIP INTERACTS SPECIFICALLY WITH THE BRCT DOMAINS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN OTIC VESICLES AT
CC DAY 9.5. AT DAY 10.5, THIS EXPRESSION DECREASES AND HIGH LEVELS
CC ARE FOUND IN THE NEUROECTODERM. AT DAYS 11-12.5, HIGH LEVELS IN
CC DIFFERENTIATING KERATINOCYTES AND WHISKER PAD PRIMORDIA. AT DAYS
CC 14-17, EXPRESSION ALSO OBSERVED IN KIDNEY EPITHELIAL CELLS. IN
CC THE ADULT, HIGHEST LEVELS FOUND IN SPLEEN, THYMUS, LYMPH NODES,
CC EPITHELIAL ORGANS, AND ALVEOLAR AND DUCTAL EPITHELIAL CELLS OF
CC THE MAMMARY GLAND. VERY LOW LEVELS IN BRAIN, KIDNEY, AND SKIN. NO
CC EXPRESSION IN HEART, LIVER OR LUNG.
CC -!- DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES
CC DRAMATICALLY DURING PREGNANCY. LEVELS FALL DURING LACTATION AND
CC INCREASE AGAIN DURING POST-LACTATIONAL REGRESSION OF THE
CC MAMMARY GLAND.
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC -!- SIMILARITY: CONTAINS 2 BRCT DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U35641; AAB17113.1; -
DR EMBL; U31625; AAB17114.1; -
DR EMBL; U32446; AAA96393.1; -
DR EMBL; U36475; AAC52323.1; -
DR EMBL; U33835; AAA99742.1; -
DR MGD; MGI:104537; Brcal.
DR InterPro; IPR001357; -
DR InterPro; IPR001841; -
DR InterPro; IPR002378; -
DR Pfam; PF00533; BRC1; 2.
DR Pfam; PF00097; zf-C3HC4; 1.
DR PRINTS; PR00493; BRSTCANCER1.
DR PROSITE; PS00172; BRC1; 2.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
KW zinc-finger; DNA-binding; Nuclear protein; Anti-oncogene; Repeat.
FT ZN_FING 24 64
FT DOMAIN 1585 1679
FT DOMAIN 1698 1797
FT VARIANT 93 93
FT VARIANT 305 305
FT VARIANT 319 319
FT VARIANT 377 377
FT VARIANT 550 550
FT VARIANT 652 652
FT VARIANT 765 765
FT VARIANT 917 917
FT VARIANT 933 933
FT VARIANT 1122 1122
FT VARIANT 1206 1206
FT VARIANT 1212 1213
FT VARIANT 1255 1255
FT VARIANT 1261 1261
FT VARIANT 1264 1264
FT VARIANT 1269 1269
FT VARIANT 1283 1283
FT VARIANT 1337 1337
FT VARIANT 1349 1349
FT VARIANT 1352 1353
FT VARIANT 1381 1381
FT VARIANT 1390 1390
FT VARIANT 1400 1400
FT VARIANT 1503 1503
FT VARIANT 1549 1549
FT VARIANT 1680 1680
FT VARIANT 1712 1712
FT VARIANT 1721 1721
SQ SEQUENCE 1812 AA; 198669 MW; 2291EA74150BB86A CRC64;

Query Match 2.1%; Score 192.5; DB 1; Length 1812;
Best Local Similarity 18.2%; Pred. No. 0.025;
Matches 260; Conservative 178; Mismatches 518; Indels 471; Gaps 64;
QY 27 LEFDLNTKTPQKPSKR--KRKEMPVVVEGKP-KRKPRK--AEL----- 66
DB 588 LEVELNVHSSKAPKNRLRRKSSIRCALPLEISRNPSPTCAELQIDSCGSSEETKKNH 647
QY 67 -----PKVVEGPKPKPRKRAAQEKVKSKETGSAKKNLKSATKKPANVGM 115
DB 648 SNOQAAGHLREPQLIEDTEPAADAKKNPEHNRKRASDAFPEE-----KL 694
QY 116 SKNSPEVTLKSC---RKALFNLEPNQDARQDSSSEIVQNSGANSFSEIRDAIGTNG 172
DB 695 MNKAGLLT--SCSSPRKSG--PVNPSPORTGTEQLETRQMSDSA---KELGDRVLGGEP 747
QY 173 SFLDVSQIDKTNGLGAMNQPLEVSGNQPDKL--STGAKLARDQOPDLITRNQOCQFPV 230

DB 748 S-----GKTTDRSEESTSVSLVSDTDYD----- 770
QY 231 ATQNTQFPMENQQAAMLMKNQLIGFPFQNGQPRMTIRNQOQCLAMGNOQPM-YLIGTPRP 289
DB 771 -TONSVSVLDAH-----TVRYAR-----TGSACQMTQFVASENP 803
QY 290 -ALVSGNQOLG-GPOGNKRPIFLNHQTCPLPAGNQLYSGPTDMHQLVMSTGGQOHLILKN 347
DB 804 KELVHGSNNAGSGTEGLKPP--LRHAL-----NLSEKVEMEDSELDQYQLNTFOVSK 855
QY 348 QQPGSLI---RGOQPCVPLIDQOPATPKGFTHLNQWATSMSSPCLRPHSQSQVPTTVL 403
DB 856 ROSFALFSKPRSPQKDCAHVSPSKELSPK-----VTAKGKQKROGQEEFEIS 903
QY 404 HVESVSRILNGTTGTQCRSRAPAYDSLOQDIHQGNKYI--LSHEISNGNCG-----KKALPQ 458
DB 904 HYQAVAAIV-GLPVPQCEGLAA--DTM---CDRGCLCPSSHYSRGENGLSATKSGISQ 958
QY 459 N-----SSLPTPIMAKLEARGSKROYHRAMGOTERKHDNLALAOOIAOSQOVRH 507
DB 959 NSHFQSVSPIRSSIKTDNRKPLTEGR-----FERR 989
QY 508 NSSTCEVYLDAAKKTKIQKVQENLHGMPEVIEIEDDPTDGARKGKNTASISKASKGN 567
DB 990 TSST-----EMAVGNENILQSTVH-----TVSLNHRGNACQAGSGSHEVCSTGD 1035
QY 568 SSPVKKTAEKEKCIIVPKTPAKKGRAGRKSVPPPAHASEIQLWQTPPKT-----PLSRSK 623
DB 1036 SFP-----GQLGRNRG-----PKVNVPPLDMSQ 1059
QY 624 PKGKRKSIQDSGK---ARGPSGELLQD-----SIATIRMQNLYLG 664
DB 1060 P-GVCQQSVFVSDKYLETIKQEGEAVCADFSPCLFSDHLEQSMGKVFQVCSCTPDDL 1118
QY 665 DKERE-----QEONAWLYKGDGALVPEYSEKRRPKVDIDDETTRINLLMCK 714
DB 1119 DVEKQHTSFGEGDIMERSAVF-----NGSILRES--SRSPSPVTHASKSQS-----LHRA 1168
QY 715 GDEKEGDEEKKKKKWEERRVFR-----GRADSFIAHMLVQGRDRFSPW 762
DB 1169 SRKLESSEESDSTE---DEDLPCFQHLLSRISNTPELTCRCSSAVTQRMPEKAGTOAPW 1224
QY 763 KGSVDSVIGVFLTONVSDHLSAPMSLAARFPKLLSSREDERVRSVYVVEDPCCIL 822
DB 1225 KGSDDCNNEVINIEASQEHQFSE-----DPRCSGMFSQSH-SAVQGSTPNANSQ 1274
QY 823 NLNEIPSMQEKVQHPDSMEVSGVDSGSKQLRDCNSGIERFNFLEKSTONLEEEVLSSQ 882
DB 1275 DSNFIPP-----STQSHQCGN-----EAFSL-- 1297
QY 883 DSFDPALFQSCGRVSGSCSKSDAEFPPTTRCETKTVSGTSQVQSPNLSDEICLOQNE 942
DB 1298 ---DKELISDNEEMATCLEEDNDQE-----EDSIIPDSEASGYESETTLESD-CSQSDI 1347
QY 943 RPHLYEGSDVQKQETTVAQKKPDLEKTMNWKDSVCFQPRNDTNWQTPPSSVEOCAT 1002
DB 1348 LP---TQEGATKYNLIKLOQEMAHLEAVLEQGNQSSGH-----SPSLGDPICAL 1395
QY 1003 RQPHVLDIEDFCMOGEGGLGYSNMSISPRVDRVKNKNVPRFRQ----- 1046
DB 1396 EDLPVLEP---NWSGAAI-----LTSKNINENPVSONLKSACDDKFOLOHLE 1439
QY 1047 GGSVPREFTGOIIPSTPHELPGMLGSSSSVAQEOHQDDTQHQQDEMKNKASHLQKTFDL 1106
DB 1440 GFTSGDDESGMGRPS-PFKSPLAGSRGSAHCSRHQLKRNPSQBELLPAG----- 1490
QY 1107 LNSSECLTRQSTQKIT-DCCLPRDRTAEDVDPDLNNSLQNLILVESNSNKEQTAV 1165
DB 1491 -----SEASSEPHNSTGESCLPR-RELEG--TPYLGSGISLFSRDPSESPKPAHI 1540
QY 1166 EYKETNATILREMKGTFLADGKKPTSQW-----SLRKQDVEGNEGROERNKNDS-- 1215
DB 1541 GTTPASTSVLKIPQGOVAFRSAAAGKAVVGIIVSKIKPELTSTSEERADRIDISWVUSGL 1600

Qy 1216 IDYFAIRRASISE-----ISEAIXERGMNNMLAVRIKDFLERIVKDHGGIDLEWLRSP 1270
Dy 1601 TPKEVMTVKFAERYLTLTDAITEETHVIKTKDAEVCERTLK-----1645
Qy 1271 DKADYLLSING-----IGLSKSEVCVRLTLTHNLAFFVDVTNVR 1309
Dy 1646 -----YFLGIAGGKWIYSVSWVRSIQERRLLNHEFEVTDGVDVTGR 1687

RESULT 12

CBP_HUMAN
ID CBP_HUMAN STANDARD; PRT; 2442 AA.
AC Q92793; Q16376; O00147;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CREB-BINDING PROTEIN.
GN CREBBP OR CBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97385172; PubMed=9238046;
RA Sobulo O.M., Borow J., Tomek R., Reshimi S., Harden A.,
RA Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
RA Zelenik-Le N.J.;
RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
RT acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=97321049; PubMed=9177780;
RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
RA Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA Doggett N.A., Peters D.J.M., Breuning M.H.;
RT "Construction of a 1.2-Mb contig surrounding, and molecular analysis
RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT 16p13.3.";
RL Genomics 42:96-144(1997).
[3]
RN SEQUENCE OF 1-405 FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borow J., Stanton V.P., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
RA Housman D., Miteiman F., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC T(8;16)(P11,P13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)
CC INVOLVING CBP AND MLL.
CC -1- DISEASE: DEFECTS IN CBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
CC SYNDROME (RUBS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
CC ABNORMALITIES, BROAD THUMBES, BROAD BIG TOES, MENTAL RETARDATION
CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U47741; AAC51770.1; -;
DR EMBL; U85962; AAC51331.1; -;
DR EMBL; U89354; AAC51339.1; -;
DR EMBL; U89355; AAC51340.1; -;
DR MIM; 600140; -;
DR InterPro; IPR000197; -;
DR InterPro; IPR000433; -;
DR InterPro; IPR001487; -;
DR Pfam; PF00569; Z2; 1;
DR Pfam; PF00439; bromodomain; 1;
DR Pfam; PF02135; zf-TAZ; 2;
DR PRINTS; PR00503; BROMODOMAIN.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00633; BROMODOMAIN_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW Chromosomal translocation.
FT DOMAIN 363 430 CYS/HIS-RICH.
FT DOMAIN 452 683 CREB-BINDING.
FT DOMAIN 1103 1175 BROMODOMAIN.
FT DOMAIN 1061 1064 POLY-GLU.
FT DOMAIN 1199 1487 CYS/HIS-RICH.
FT DOMAIN 1555 1562 POLY-GLU.
FT DOMAIN 1675 1849 CYS/HIS-RICH.
FT DOMAIN 1943 1948 POLY-PRO.
FT DOMAIN 2081 2085 POLY-GLN.
FT DOMAIN 2199 2216 POLY-GLN.
FT DOMAIN 2245 2248 POLY-GLN.
FT DOMAIN 2297 2300 POLY-GLN.
FT CONFLICT 1511 1513 FAE -> NSG (IN REF. 2).
FT CONFLICT 1724 1725 ED -> VV (IN REF. 2).
FT CONFLICT 1770 1770 V -> L (IN REF. 2).
FT CONFLICT 1789 1789 N -> F (IN REF. 2).
FT CONFLICT 1812 1812 T -> P (IN REF. 2).
SQ SEQUENCE 2442 AA; 265336 MW; 42D084619475F3D2 CRC64;

Query Match 2.1%; Score 192; DB 1; Length 2442;

Best Local Similarity 18.1%; Pred. No. 0.04;
Matches 345; Conservative 238; Mismatches 720; Indels 598; Gaps 89;

Qy 5 MDSAVNATEATEONGSDROVLEFDLNT-----POKPSKRRKRPMPKVVVGKPKRK 59
Dy 288 MGATGVNPQLASKQSMVNSLPTFTDIKNTSVTPNNSQMOTSVGIVPTQAIATGTAD 347
Qy 60 P--RKPALPKVVVGKPKRPKRAATQEKVKSKETGSAKKKLNKESATK----KPAVNG 113
Dy 348 PERKRLIQOQLVLLHAKCORREQANGEVRACSLPHCRMTMKNVNLNHTHCQAGKACQVA 407
Qy 114 DMSNKSPEVT-LKSCRK---ALNFDLENPGDARQD-----SESEIVQSSGSAFSFIRD 165
Dy 408 HCASSRQIISHWKNCTRHDCPVCLPLKNASDKRNQQTILGSPASGIQNTIGS-----V 460
Qy 166 AIGTNGSFSDVSQIDKTN-----GLGAMNQPLEV-----196
Dy 461 GTGQGNATSLNPNPIDSSNQRAYAALGLPYMNPOTQLQPVPVGGQPAQPOTHOQMRT 520
Qy 197 --SNGNOPDLSTCAKLARDQOQDLTRNQOCOPPVATONTQPEMNOQAWLQMKNLIG 254
Dy 521 LNPLGNPMNIPAGG-ITTDQPPNLLISEALPTSIGATN---PLMND-----G 565
Qy 255 FPFQGNQPRMTIRNQQLCLAMGNOQPMY-----LIGTPRPA-----290
Dy 566 SNSGNGTSTLTPTAAPPSTSGVRKGWHEVTDLRSHLVHKLVAQIFPTDPAAKLDKR 625
Qy 291 ---LVSGNQQLGG---PQGNKRPIFLNHQCLCPAGNOLYGSPTDMHOLVMSGTGQHQGLL 344
Dy 626 MENLVAYAKVEGDMYSEANSRDEYH-----LLAEKIYKIQKELEEKRRS---RLKQG 677
Qy 345 IKNOQPSGLIRGOQPCVPLDQ-OPATPKGFTHLNQMVATSMSSPGLRPHRSQOVPTYL 403
Dy 678 ILGNQPALPAPGAQP--PVIQAQVVRP-----PNGSLSLPVNRM 715

QY 404 HVESYRIILG-TTGTGCRSAPAYDSLOODIHQGNKYLISHSISNGNGCKKALPQNSSL 462
 Db 716 QVSOQMGNSFNMSLGNVQLPOAPM-----GPRASPMNHVS 751
 QY 463 PTPIMAKLEEARGSKRQYHRAMGQTEKHDLN-LAQOIAQSQDVERHNSSTVCVEYLDA--- 518
 Db 752 QMNSMGVPGMAISPRMPQPPNMGGAHTNNMMAQAPASQFLPQNPQFPSSGAMSVGMG 811
 QY 519 -----AKTKIKQVVOENLH----- 533
 Db 812 OPPAQTVSOGVPGAALPNLMLGPOASQLPCPPVTSQPLHPTPPASTAAGMPSLQH 871
 QY 534 ---GMPPEVIEIDPTDGARKGNTASISKG-----ASKGNSSP-VKKTAEKEKCIYVKP 584
 Db 872 TTPQMTTPQPAAPTQPSVSSGQTPPTPGSVPSATQSTPTVQAAAQAQVTPQPO 931
 QY 585 TPAKKGAGRAKRS---VPPPAH-----ASEIQLWQTP-----PK 616
 Db 932 TPVQPPSVATPQSSQQQTPVHAQPPGTPLSQAAASIDNRVPTPSSVASAETNSQOQPGD 991
 QY 617 TPLSRKPKGKRKSIQDSGKARG-PSGELLQDSIAEIIYRMQNLXLYLGDKEREQEQNAM 675
 Db 992 VPVLEKMTQEAETPDGSGESKGPRESMEED-----LQASQVKEETDI 1038
 QY 676 VLYKGDGALVPYSEKRPKPRKVIDDDETRIWNLLMGKDEKDEKDKKKKXWEE 735
 Db 1039 ABQKSE---PMEVDEKPEVKEVEKEEESNN---GTASQSTSPSPQPKKIFK-PEEL 1090
 QY 736 RVFRGRADSFATRHLYVQGRDRFRSPWKGSVVDVSVIGVFLTONVSDHLSSSAFMSLAARF 795
 Db 1091 QOALMPTLEA-----LYRQDPESLPFPQDPQLLGI---PDYDIVKNPMDLSTIKR- 1140
 QY 796 PPKLSSREDE--RNRSVVVDPPGCIILN--LNEIPGQKRVQHPDSMEVS-----GV 845
 Db 1141 --KLDGTQYQEPQYQVDDVLMFNNAWLYNRKTSVYKFCSLAEVFEQIEDPVMQSLGY 1198
 QY 846 DSGSK-----EQL-----RDCNSGTE-RFNLEKSIQNLSEVLS----- 880
 Db 1199 CGGRKYEFPQTLCCYQKQOLCTIPDAAYSYQNRHFECEKFTTEIQGENVFLGDDPSQ 1258
 QY 881 -----SDPSFPAIFQSGRGVS-----C-SCRSKDAEF 908
 Db 1259 QTTISKQDFEKKKNTLDPEPPVDCCKGRKMHQICVLHYDIWPSGFCVDCNCKTKGR- 1317
 QY 909 PTTRETCTVSTQSQTGSPNLSDIEI--CLOGNERPHLYEGSDV-----QKQETTN 960
 Db 1318 --PRKENFSAKRLQTLRLGN-HLEDVRNKFELRRQNHPE-----AGEVFVRVVASDKTVE 1370
 QY 961 VQAKKPDLEKTMNWKDVCFCQPRNDTNQWTPSSSYEQCATRQPHVLDIEDFGWQEGCL 1020
 Db 1371 V---KPGMKRS--RFVDS---GEMSESFPYRTKALFAFEEI-----DGVDVCFGMHVQY 1417
 QY 1021 GYSWMSISPRVDRVKNKNVPR-RFRQGSVPR-----EF-----TGO 1057
 Db 1418 GSD--CPPNTRRVISYLDSTHFR-----PRLTAVYHILLIGLEYVKKLGVYVGH 1470
 QY 1058 ITPSTPHLPGLMGSGSSAVQEHQDDTQHNQODDMKASHLQKTFDLNLNS--SEECIT 1115
 Db 1471 IWACPSE-----GDDYIFHCHPPD-----QKIPKPKLQENYKMKLDKAFERI 1517
 QY 1116 ROSSYKQNTDCLPRDRTAEDVDVPLNNSSLQNLIVESNNSKEQTAVEYKFNATIL 1175
 Db 1518 DYKIDFKQATE-----DRUTSAKELPYEGDPWPVNLVEESIKELEQEEERKESTAS 1572
 QY 1176 REMKGTADGKKPTQOWDSLRKDVEGNEGROERNKNMDSIDYAIRASISEISEAKE 1235
 Db 1573 ETTEGSGDSK-----NAKKKNKTKNNK-----SSIRANKK 1606
 QY 1236 R---GMNNLAVRIKDFLER-----IVKHGGIDLEWLRESPPDKADYLLISIRGLGL 1285
 Db 1607 KFSMPNVSNLQKLYATMEKHEVFFVHLHAGPVINTL---PPIVDPDPLLS----- 1657
 QY 1286 KSVCEVRLTLNLNAPVDNTNVR-----IAVRMGWVLPQLPESLQLHLLELYPVLES 1340

RESULT 13

TPR_HUMAN ID TPR_HUMAN STANDARD; PRT; 2349 AA.
 AC P12270;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NUCLEOPROTEIN TPR.
 GN TPR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93064711; PubMed=1437155;
 RA Mitchell P.J., Cooper C.S.;
 RT "The human tpr gene encodes a protein of 2094 amino acids that has
 extensive coiled-coil regions and an acidic C-terminal domain.";
 RL Oncogene 7:2329-2333(1992).
 RN [2]
 RP REVISIONS, AND CHARACTERIZATION.
 RX MEDLINE=95096166; PubMed=7798308;
 RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
 Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
 RT "tpr, a large coiled coil protein whose amino terminus is involved in
 activation of oncogenic kinases, is localized to the cytoplasmic
 surface of the nuclear pore complex.";
 RL J. Cell Biol. 127:1515-1526(1994).
 RN [3]
 RP SEQUENCE OF 1-142 FROM N.A.
 RX MEDLINE=88262257; PubMed=3387099;
 RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
 RT "tpr homologues activate met and raf.";
 RL Oncogene 2:617-619(1988).
 CC -!- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
 COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
 IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
 COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
 TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
 COMPONENTS, INCLUDING P62.
 CC -!- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
 BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
 CC -!- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
 OR RAF GENES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

QY 1517 LVALHPTTTSIPTKLNKNSIRLTERHGVYELPDHR-----LLDGMKREPDSP 1567
 DB 1219 QVAVTPP-----PKETPOISEYQYSELYSANSKNVSELAYLDDLDKL--DDIDE 1272
 QY 1568 YLLAIWTPGET 1578
 DB 1273 YLLNNIMPEKT 1283

RESULT 15

MAPB_MOUSE STANDARD; PRT; 2464 AA.
 AC P14873;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MICROBULE-ASSOCIATED PROTEIN 1B (MAP1.2) (MAP1(X)) [CONTAINS: MAP1
 DE LIGHT CHAIN LCL1.
 GN MAP1B OR MTAP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RX MEDLINE=90094539; PubMed=2480963;
 RA Noble M., Lewis S.A., Cowan N.J.;
 RT "The microtubule binding domain of microtubule-associated protein
 RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
 RT and tau".
 RL J. Cell Biol. 109:3367-3376(1989).
 CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
 CC STABILIZING MICROTUBULES.
 CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LCL1, LCL2 AND LCL3, CAN ASSOCIATE
 CC WITH MAP1A AND MAP1B PROTEINS.
 CC -!- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE
 CC KKEE AND KKEI/V. REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER
 CC REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES
 CC BOTH IN VITRO AND IN VIVO.
 CC -!- PTM: LCL1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
 CC OF MAP1B.
 CC -!- SIMILARITY: TO NEURAXIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. Usage by and for commercial
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X51396; CAA35761.1; -;
 CC PIR; S07549; QRMSP1.
 CC MGD; MGI:97179; Mtap5.
 CC InterPro; IPR000102; -;
 CC Pfam; PF00414; MAP1B_neuraxin; 10.
 CC PROSITE; PS00230; MAP1B_NEURAXIN; 7.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2464 MAP1 LIGHT CHAIN LCL1.
 FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 FT KKEE AND KKEI/V REPEATS).
 FT DOMAIN 1865 2068 12 X 17 AA TANDEM REPEATS.
 FT REPEAT 1865 1881 1.
 FT REPEAT 1882 1898 2.
 FT REPEAT 1899 1915 3.

FT REPEAT 1916 1932 4.
 FT REPEAT 1933 1949 5.
 FT REPEAT 1950 1966 6.
 FT REPEAT 1967 1983 7.
 FT REPEAT 1984 2000 8.
 FT REPEAT 2001 2017 9.
 FT REPEAT 2018 2034 10.
 FT REPEAT 2035 2051 11.
 FT REPEAT 2052 2068 12.
 SO SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDBDA87 CRC64;

Query Match 2.1%; Score 190; DB 1; Length 2464;
 Best Local Similarity 18.6%; Pred. No. 0.051;
 Matches 291; Conservative 226; Mismatches 640; Indels 410; Gaps 67;

QY 175 LDSVSQIDKT-----NLGAMNQPLEVSMGNQPKLSTGAKLARDQDPDLTRNQOCQPPV 230
 DB 301 LDRVDSILLTHIGDDNLPGLNSMLQRIAELEERSOGSTNSDMKMLISPDGLGVVFLN 360
 QY 231 ATONTOPFMEQOAWLQMKNLQIGFPGNQOPMTIRNQOPL-AMGNO-OPMYL---IG 285
 DB 361 VPENLKDPEN-----IRMKRSIEEACFTLQYLNKLSMKPEPLFRSVGNTIEPVILFKMG 416
 QY 286 TPRPALVSGN-----QOLGGPOGNKRPFLFNHQTCLPAGNQL---YGSPTDMH 330
 DB 417 VGKLEMYVLPVSKSKEMQYFMQQTGINKDAELI-----LPGQEVDPISVLTSSV 470
 QY 331 QLVNSTGQOQHLLIKNOQP-----SLIRG-----QPCV---PLIDQOQATPK 372
 DB 471 SLIYVHPANPAEKIIRVLPFGNSTQYNIIEGLEKHLKHLDFLQPLATQKOLTGQVTPPV 530
 QY 373 GFTHLNOMV-----ATS-MSSPGLRPHSQSOVPTTYLHVSVSRILNCTGTGTCORSR 423
 DB 531 KOVLKORADRESLKPATKPVASKSVKSKETPE-----VTTTSQVEK 576
 QY 424 APAYDSQQDIDHOGNKYILSHEISNGCKKALPONSLSPTTIPMAKLEARGS-----KRQ 479
 DB 577 TPKVESKEKVLVKDKPVKTE--SKPSVTEKESVSKKE--QSPVKAQVAEKQATESKPKVT 633
 QY 480 YHRAMGQTEKHDNLNAQOIAQSDOVRHNSSTCYEYLDAAKTKIOKVVOENLHGMPEV 539
 DB 634 KQVYVKEIKTKLEEKKEPKKPKKEDTKPLKDEKPKKEVKEIKK-----684
 QY 540 IEIEDDDPTDGAKNKNTASISKGASKGNSPVKTAKEKCIPTAKKGRAGRKSVSP 599
 DB 685 -EIKKEERKELKVEKETPLDKAKVEKKEKVEKKEE-KEPKKELKLSKDIKSTP 742
 QY 600 -----PPAHASEIQLWQPTPTPLSRKPKGKR-KSIODSGKARGPSGELLCQDSIA 652
 DB 743 QSDTKKPSALKPKVAKKEESTKKEPLAAGKLKDKGVKVIKKEGKTEAATAVGTAAAT 802
 QY 653 EIIYRMQNLILGDKEREQEQNAMVLYKGDGALVYESKRRKPRKPVDDIDDETRINWLLM 712
 DB 803 AAVVAAAGIAASGVPVKELEASRLMSSPEDLTDKFEELKAE---EIDVAKDIKPOLELI- 858
 QY 713 GKGEKEDGDEKDKKKKWKWEEERRRFRGRADSIARHMLVQGDRRFRSPKWSGVDSVIG 772
 DB 859 -----EDEKLEKTOP-----GEA-----YVLOKETEVS--KGSAESDEG 892
 QY 773 VLTQNVSDHLSSSAFMSLAARFPFKLSSSRREDNRVRSVVVEDEPGCILNLEIPSQWE 832
 DB 893 ITTTEG-----EG-----E 901
 QY 833 KVQHPSDM---EYSGVDSGSKQELRDCNSGIERNFLEKSIONLEEVLSQDSFDPAI 889
 DB 902 CEQTEPELEPVKOGVDD-----IEKF-----EDEGAGFEESSETGD 938
 QY 890 FOSGCRVSGSCSDAEFFPTTRCTKTVSTGTSVOTGSPNLSDLEICLOGNERPHYEG 949
 DB 939 YEE-----KAETEAEAP-----EEDGEDNAGSASAKHSPTEDDE-SAKAEADVHLKEK 986
 QY 950 -----SGDVQKQETTNVAQKKPDLEKTMNMKDSVCFQGPORNDTNWQTTPTSSSYEQCATR- 1003

```

Db 987 RESVVGDDRAEDMDVLEKEAQSEE-----EGEEDKADAREEGYEPDKTEA 1038
QY 1004 QPHVLIED-----FGMGEGGLGYSMMSISPRVDRVKNKNVPRFRFQGG--SVPREFTGQ 1057
Db 1039 EDYVMAVADKAAAGVTEQYGLTSA-----KPGIQSPSREPASS 1081
QY 1058 IIPSPHPELPGMGLSGSSSAVOEHODDTHOQODENKASHLOKTFLLDLNSESBECLTRQ 1117
Db 1082 IHDET---LPG---GSESEATASDENEDQPEFTATSGYTQSTIEI--SSEPTPMD 1132
QY 1118 SSTKONITDGCCLPRDRTAEDVDPLSNSSNLILVESNNSNKEOTAVEYKETNATILRE 1177
Db 1133 MST-----PR-----DMSDETNEETESPQOEFVITKYESSLYSQEYSKPAVAS 1178
QY 1178 MKGTLAGKKPTSQWDSLRKQVNEGGR--OERNKNMDSIDYEATRASIIEISEIAIKER 1236
Db 1179 FNG-LSSEGSK-TDATDG--KDYNASASTISPPSMEEDKFSKALRDAYCSE-----EK 1228
QY 1237 GNMNLAIVRIKDFLERIVKDHGIDLEWLRSPDPKADYLLSIRGLGLKSVCEVRLTL 1296
Db 1229 ELKASAEIDIKOVSDERLSPAKSPSLSPSPPIEKP-----LGRSVNF--SLTP 1278
QY 1297 HNLAPPVDTNVGRIAVRMGWVPLQPLPESLQHLHLLLYPVLESIOKFLWPRCLKLDQRTL 1356
Db 1279 NEIKVSAEGEARSVSGVGTQAVVEHCASPEEKTLEVVSPSQSV-----TGSAG 1327
QY 1357 YELHYQLITFGKV--FCTKSRPNCNACMRGECRHFASAYASARLALPAPEERSLTSATI 1414
Db 1328 HTPYQSPTEDEKSSHLPTVSENAQVPYSFE--FSEAKDENERASLSPMDEPVPDS 1384
QY 1415 PVPDESFPVPAITMIELPLPLEKSLA-----SGAPSNRENCEPIIEEPASQOECT 1465
Db 1385 PVEKVLPLRSPPLLGSESPYEDFLSADSKVLGRRSESPFEGKNGKQGPDPRESVSDLT 1444
QY 1466 -----EITESDIEDAYNE-----DPDEIPIKINI 1491
Db 1445 STGLYQDKOEKSTGFIPIKEDFGPEKTSVDVTWSSQSALALDERKLGCDVSPT-OIDV 1503
QY 1492 EQFGMTLREHMERNMELQEG-----DMSKALVALHPTTT 1525
Db 1504 SQFG-SFKE--DTKMSISEGTVSDKSATPVDEGVAEDTYSHMEGVASVSTASVA---TS 1556
QY 1526 SIPTPKLKNISRLTEHQYVELPDSHRLLDGMDKREPDPDPSYLLAIWTP---GETANSA 1582
Db 1557 SFPEPTDDVS-----PSLHAEVGSPHSTEVDD-SLSVSVVQPTTTFOETEMS- 1603
QY 1583 QPPEOKC 1589
Db 1604 -PSKEEC 1609

```

Search completed: July 5, 2001, 12:46:19
Job time: 241 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:18 ; Search time 36.19 Seconds
(without alignments)
6320.956 Million cell updates/sec

Title: PCT-US01-13059-2

Perfect score: 9089

Sequence: 1 MQSIMDSSAVNATEATEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_TREMBL_16.*

2: SP_Archea.*

3: SP_Bacteria.*

4: SP_Fungi.*

5: SP_Invertebrate.*

6: SP_Mammal.*

7: SP_MHC.*

8: SP_Organelle.*

9: SP_Phage.*

10: SP_Plant.*

11: SP_Rodent.*

12: SP_Unclassified.*

13: SP_Vertebrate.*

14: SP_Virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3949	43.4	1017	Q9L269	Q9L269 arabidopsis
2	2863	31.5	555	Q9L568	Q9L568 arabidopsis
3	1776.5	19.5	1309	Q9SR66	Q9SR66 arabidopsis
4	1559.5	17.2	1207	Q9SJQ6	Q9SJQ6 arabidopsis
5	1260	13.9	234	Q9L267	Q9L267 arabidopsis
6	999	11.0	274	Q9FTQ2	Q9FTQ2 oryza sativ
7	709	7.8	917	Q9498	Q9498 arabidopsis
8	281	3.1	2151	Q9NG79	Q9NG79 trichomonas
9	278	3.1	5327	Q9T691	Q9T691 drosophila
10	245	2.7	2254	Q9HCY0	Q9HCY0 homo sapien
11	241	2.7	2187	Q9H197	Q9H197 homo sapien
12	240.5	2.6	2938	Q9L769	Q9L769 mus musculus
13	239	2.6	2701	Q9Y520	Q9Y520 homo sapien
14	238	2.6	5120	Q9PU36	Q9PU36 gallus gall
15	235.5	2.6	4880	Q9JLT1	Q9JLT1 rattus norv
16	235.5	2.6	5085	Q9JKS6	Q9JKS6 rattus norv
17	233.5	2.6	2218	Q9SL26	Q9SL26 arabidopsis
18	232	2.6	259	Q9RRQ0	Q9RRQ0 deinococcus
19	232	2.6	2290	Q9VT00	Q9VT00 drosophila

ALIGNMENTS

RESULT 1

Q9L269

ID Q9L269 PRELIMINARY; PRT; 1017 AA.

AC Q9L269;

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)

DE HYPOTHETICAL 112.1 KDA PROTEIN.

GN T32M21_160.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Terryn N., Ardiles W., Buysshaert C., Dasseville R.,

RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,

RA Villarroel R., Gieles J., Van Montagu M., Bancroft I., Mewes H.W.,

RA Rudd S., Lemcke K., Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project.

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL162875; CAB85562.1; -

KW Hypothetical protein.

SQ SEQUENCE 1017 AA; 112139 MW; 4F7C313A891EC83F CRC64;

Query Match 43.4%; Score 3949; DB 10; Length 1017;
Best Local Similarity 99.9%; Pred. No. 8.2e-257;
Matches 755; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSIMDSSAVNATEATEQNDGSRQDVLEFDLNKTPQKPSKRRKFKMPKVVVEGPKRRKP 60

DB 235 MQSIMDSSAVNATEATEQNDGSRQDVLEFDLNKTPQKPSKRRKFKMPKVVVEGPKRRKP 294

QY 61 RKAPELPKVVVEGPKRRKPRKAATQEKVSKETGSAKKKNLKAESATKKPANYGDMNSKSP 120

DB 295 RKAPELPKVVVEGPKRRKPRKAATQEKVSKETGSAKKKNLKAESATKKPANYGDMNSKSP 354

QY 121 EVTLKSCRKALNFLENPGDARQDSESEIVQNSGANSFSEIRDAIGTGNGSFLDSVSQ 180
DB 355 EVTLKSCRKALNFLENPGDARQDSESEIVQNSGANSFSEIRDAIGTGNGSFLDSVSQ 414
QY 181 IDKTNGLGAMNOLPVEVNGNODKLSLTKAGAKLARDQDPDLTRNOCCOPFPVATQNTQPM 240
DB 415 IDKTNGLGAMNOLPVEVNGNODKLSLTKAGAKLARDQDPDLTRNOCCOPFPVATQNTQPM 474
QY 241 NOQAWLQMNOLIGFPNGNOQPMRTIRNOQPCLAGMNOQPMYLGTPRALVSNQOOLGG 300
DB 475 NOQAWLQMNOLIGFPNGNOQPMRTIRNOQPCLAGMNOQPMYLGTPRALVSNQOOLGG 534
QY 301 PQGNKRPFILNHOQCLPAGNOLYGSPTDMHQLVMSTGGQGHLLIKNQOPGSLIRGOQPC 360
DB 535 PQGNKRPFILNHOQCLPAGNOLYGSPTDMHQLVMSTGGQGHLLIKNQOPGSLIRGOQPC 594
QY 361 VPLIDQOPATPKGTHLNQMVATSNMSPGLRPHSQSOVPTTYLHVESVSRLLNGTTCTCQ 420
DB 595 VPLIDQOPATPKGTHLNQMVATSNMSPGLRPHSQSOVPTTYLHVESVSRLLNGTTCTCQ 654
QY 421 RSRAPAVDSLOQDTHQGNKYILSHSINGNGCKKALPQNSLSLPTPIMAKLEEAARGSKROY 480
DB 655 RSRAPAVDSLOQDTHQGNKYILSHSINGNGCKKALPQNSLSLPTPIMAKLEEAARGSKROY 714
QY 481 HRAMGOTEKHDLNLAQOIAQSDQVERHNSSTCVYELDAAKTKIKQVQENLHGMPEVI 540
DB 715 HRAMGOTEKHDLNLAQOIAQSDQVERHNSSTCVYELDAAKTKIKQVQENLHGMPEVI 774
QY 541 EIEDDPTDGARKGNTASISKGASKGNSPVKTAEEKICIVPKTAKGRAGRKSVPP 600
DB 775 EIEDDPTDGARKGNTASISKGASKGNSPVKTAEEKICIVPKTAKGRAGRKSVPP 834
QY 601 PAHASETQLOMPTPKPLSRKPKGKRSIQDSGRARGSPGELLCCQDSTAEIIRYMON 660
DB 835 PAHASETQLOMPTPKPLSRKPKGKRSIQDSGRARGSPGELLCCQDSTAEIIRYMON 894
QY 661 LYLGDKEREQEQANAVLYKGDGALVPYESKRRKPRKPRVDIDDETRITWNLIMGKDEKEG 720
DB 895 LYLGDKEREQEQANAVLYKGDGALVPYESKRRKPRKPRVDIDDETRITWNLIMGKDEKEG 954
QY 721 DEEKDKKKKKWEEERRVFRGRADSFARMHLVOGD 756
DB 955 DEEKDKKKKKWEEERRVFRGRADSFARMHLVOGE 990

RESULT 2

Q9LZ68 PRELIMINARY; PRT; 555 AA.
AC Q9LZ68;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 62.1 KDA PROTEIN.
GN T32M21.170.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villaroel R., Gleen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL162875; CAB85563.1;
DR InterPro: IPR003265;
DR SMART: SM00478; ENDO3c; 1.

KW Hypothetical protein.
SQ SEQUENCE 555 AA; 62088 MW; A1B44BDDEL17FDC1E CRC64;
Query Match 31.5%; Score 2863; DB 10; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.4e-184;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 785 SSAPMSLAARFPKPLSSRREDERNVRVSVVDEDPGCTILNLEIPSWOEKVOHPSDMEVSG 844
DB 7 SSAPMSLAARFPKPLSSRREDERNVRVSVVDEDPGCTILNLEIPSWOEKVOHPSDMEVSG 66
QY 845 VDSGSKQLRDCNSGIERENFLEKSIQNLLEEELSSQDSFDPALFOSCGRVGSCSCSKS 904
DB 67 VDSGSKQLRDCNSGIERENFLEKSIQNLLEEELSSQDSFDPALFOSCGRVGSCSCSKS 126
QY 905 DAEPTTRCTETKTVSGTSQSVQTSQSPNLSDIEICLQNERPHLYEGSGDVQKQETTNAQK 964
DB 127 DAEPTTRCTETKTVSGTSQSVQTSQSPNLSDIEICLQNERPHLYEGSGDVQKQETTNAQK 186
QY 965 KPDLKTMNKKDSVCFQGPNDTNWQTPSSSYEOCATRPHVLDIEDFGMOGEGLCYSW 1024
DB 187 KPDLKTMNKKDSVCFQGPNDTNWQTPSSSYEQCATRPHVLDIEDFGMOGEGLCYSW 246
QY 1025 MSISPRVDRVKNKVNPRFRQGSVPREFTGQIIPSTHELPGMGLSGSSSAVQEHQDD 1084
DB 247 MSISPRVDRVKNKVNPRFRQGSVPREFTGQIIPSTHELPGMGLSGSSSAVQEHQDD 306
QY 1085 TOHNOQDEMKNASHLQKTFDLNLSSECLTRQSTTKONITDGCPLPRDRTAEDVVDPLSN 1144
DB 307 TOHNOQDEMKNASHLQKTFDLNLSSECLTRQSTTKONITDGCPLPRDRTAEDVVDPLSN 366
QY 1145 NSSLQNTLVESNSNKEQTAVEYKETNATILREMKGTADGKKPTSDWDSLRKDVCEGNEG 1204
DB 367 NSSLQNTLVESNSNKEQTAVEYKETNATILREMKGTADGKKPTSDWDSLRKDVCEGNEG 426
QY 1205 RQERNKNMDSIDYEAIRRASISEISEIAKERGMNNMNAVRIKDFLERIVKDHGGIDLEW 1264
DB 427 RQERNKNMDSIDYEAIRRASISEISEIAKERGMNNMNAVRIKDFLERIVKDHGGIDLEW 486
QY 1265 LRESPDPKADYLLSIRGLGKSEVCRLLTLHLNAPVDVTNGRIAVRMGWVPLQPLPE 1324
DB 487 LRESPDPKADYLLSIRGLGKSEVCRLLTLHLNAPVDVTNGRIAVRMGWVPLQPLPE 546
QY 1325 SLQLHLEL 1333
DB 547 SLQLHLEL 555
RESULT 3
Q9SR66 PRELIMINARY; PRT; 1309 AA.
AC Q9SR66;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE T22K18.18 PROTEIN.
GN T22K18.18
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLOMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC010927; AAF04422.1;
DR InterPro: IPR003265;


```
QY 230 VATONTQPMENQAWLQMKNLQIGFPFGNQQRMTIRNOQPCLAMGNQPMYLIGTRPR 289
Db 272 -----QKHWLYSPNL-----QQGM--RYDAIC----- 293
QY 290 ALVSGNOOLGGPOGNKRPIFLNHOTCLPAGNOLYGSPTDMHOLVMSTGGQOHLIKNOQ 349
Db 294 -----SKVFSQQHNYVSFAF--HATCYSSTSQLSANRV-----LTVEERR 331
QY 350 PGSILRQOCVPLIDQOATP-----KGFTHLQMWATSMSSPGLRPHSOSQV 398
Db 332 EGIFQGRQSELANVLSKIDTPIKKTGTHARFNLSMMKLV-----EV 376
QY 399 PTVTLVHVESVRILNGTGTGCRSAPAYDSLOQDIHOGKNTILSHESINGCKKALPQ 458
Db 377 PE---HL-----TSGVCSKQO-----QNNKILVDTRVT----- 401
QY 459 NSSLPTFIMAKLEFARGSKRYHRAMQGTQEKHDNLAAQIAQODVERHNSSTCVEYLDA 518
Db 402 -----VSKKPKTKSEKQ----- 414
QY 519 AKTKIOKVQENLHGMPPVEIEDPTDGARKGKNTASISKASKGNSPVKKTAKE 578
Db 415 ---TK-QKNLLPNLCRPPSPFTGLSPD----- 437
QY 579 KCIVPKTPAKKGRAGRKSVPPPAHASEIQLWQPTPKTPLRSKPKGKGRKSIQDSGKA 638
Db 438 -----ELM-----KRNST----- 446
QY 639 RGPSEGLLODSTAETIIRYMONLYLGDKEQEONAMVLYK-----GQALVPYE 688
Db 447 -----ETISEL-----LRLLDINREHSETALVPYTMNSQIVLFGGAGAIIVPT 490
QY 689 S-KKRPRPKVDIDDETRINLLMGKDEKDEKDKKKKEKWEERERVRGRADFSI 747
Db 491 PVKPRPKVLDDETRVWKLL--ENINSEGVGDGEQAKWBERNVRFRGRADFSI 549
QY 748 ARMHLVQGRDRFSPKGSVVDVSGVIGVFLTONVSDHLSAFAFMSLAARFP--PKLSSREDE 806
Db 550 ARMHLVQGRDRFSPKGSVVDVSGVIGVFLTONVSDHLSAFAFMSLASQFPVFPSSNFD- 608
QY 807 RNVRSVVDEPEGCILNLEIPSKQ-----EKVQHPSDMEVSGVDSGSKQLRDCSNS 859
Db 609 -----AGTSSMPSIQIYLDSEETMSSPPDPDHHSSVT----- 640
QY 860 GIERENFLEKSIONLEEVLSODSDFPAIFQSGRGVSCSKSDAEFFTRCET--KT 917
Db 641 -----LKTQPDDEKDYVPSNET-----SRSSSEIAISAHESVDKT 676
QY 918 VSQTS--QSVQTSPLNLSDEICLQGNRPHLYEGSGDVOK--QETTNVAQKPKDLEKTMNW 974
Db 677 TDSKEYVDSRKG-----SVEVDKTEKCVNLNLFPSDSALTC 716
QY 975 KDSVCFQPNRDNWQTPSSSYEQCATRQPHVLDIEDFMQOEGGLYGSNMSTSPRVDRV 1034
Db 717 QHSMVSDAPONT--ERAGSS-----EIDLEGE-YRTSPMKL----- 750
QY 1035 KKNVPRFRFGSGSVPRFTGTLIPSTHELPGMGLSGSSSAVQEHODDTQHQQDEMN 1094
Db 751 -----LQGVQVLEDSNQVSPN-----MSPGDCSSEIKGFQ-----SKKEPT 787
QY 1095 KASHLQKTFDLNLSSE--CLTRQSTKQNTIDGCLPRDRTAEDVDVDPNLSLQNILV 1153
Db 788 KSS-----VDSSEPGCSQO-----DG-----DVL----- 807
QY 1154 ESNSNKEQTAVYEKTNATILREMGKTLADGKPKTSQWDSLRKDVEGNREGROERNKNNM 1213
Db 808 -----SCOKPT--LKEKGKVLKE-----EKKAFDWDCLRREAQARAGIREKTRST 852
QY 1214 DSDIYEAIRRAISTEISEAIKERMNMLAVRI-----KDFERTVVKDHGGIDLE 1263
Db 853 DTDVWKAIRAADYKVEAETIKSRGMNHKLAERTQYLTLANMKINGFGLDLRLVDHGSIDLE 912
```

```
QY 1264 WLRESPPDKADYLLSIRGLGLKSVCEVRLTLHLNLAFFPVDITNVRITAVRMGWVLPQLP 1323
Db 913 WLRDVPPDKAK----- 923
QY 1324 ESLQLHLELLEYPVLESIQKFLWPRLOKLDORTLYELHYQLITFGKVPCTKSRPNCNACPM 1383
Db 924 -----YELHYQMITFGKVPCTKSRPNCNACPM 950
QY 1384 RGCRRHFASAYASARLALPAPEERSLTSATIPVPPSPFPVPAIPMIELPLPLEKSLASGA 1443
Db 951 KGBRHFASAFASARLALPSTKGMGTDPKNPLP-----LHLPFFQREGSEV 999
QY 1444 PSNREN-----CEPIIEEPASPGQCTEITSDIEDAYVYNEDEPTIKLNIQFGMT 1497
Db 1000 VQSEPAKKVTCCEPIIEEPASPEPEAEVSIADIEEAF--EDPERIPTIRLNDAFTSN 1058
QY 1498 LRHMERNMELQGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLDGM 1557
Db 1059 LKKIMEHNKELQDGNSSALVALTAETASLPMPLKNISOLRTEHRYVELPDEHPLLAQL 1118
QY 1558 DKREPDDPSYLLAIWTPGETANSAPPEQKCGKSGKMGCFDETCSECNLSREANSQTV 1617
Db 1119 EKREPDDPSYLLAIWTPGETADSIQSVSTCIFIQANGMLCDEETCFSCNSIKETRSQIV 1178
QY 1618 RGTLLI 1623
Db 1179 RGTILV 1184
RESULT 5
Q9LZ67 PRELIMINARY; PRT; 234 AA.
AC Q9LZ67;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL, 26.6 KDA PROTEIN.
GN T32M21.180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID:3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162875; CAB85564.1; -.
KW Hypothetical protein.
SQ SEQUENCE 234 AA; 26649 MW; B395392E06EB297 CRC64;
```

Query Match 13.9%; Score 1260; DB 10; Length 234;
Best Local Similarity 100.0%; Pred. No. 5.3e-77;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1496 MTLREHWMERMELOEGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLD 1555
Db 1 MTLREHWMERMELOEGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLD 60
QY 1556 GMDKREPDDPSYLLAIWTPGETANSAPPEQKCGKSGKMGCFDETCSECNLSREANSQ 1615
Db 61 GMDKREPDDPSYLLAIWTPGETANSAPPEQKCGKSGKMGCFDETCSECNLSREANSQ 120
QY 1616 TVRGTTLLICRTAMRGSFPLNGTYTFQVNEFLFADHSSSLAPIDVPRDWINDLPRTVYFCT 1675
```



```
Db 121 TVRGILLIPCRAMRSGPLNCTYFQVNELEADHESLKPIDVPRDWLWDLPRRTVYEGT 180
Qy 1676 SVTSIFRGLSTEQIOFCFWKGFVCGVGEQKTRAPRPLMARLHFPASKLNKNT 1729
Db 181 SVTSIFRGLSTEQIOFCFWKGFVCGVGEQKTRAPRPLMARLHFPASKLNKNT 234

RESULT 6
Q9FTQ2 PRELIMINARY; PRT; 274 AA.
ID Q9FTQ2 AC
Q9FTQ2; AC
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE P0665D10.14 PROTEIN.
GN P0665D10.14
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RT Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0665D10.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002861; BAB16489.1; -.
SQ SEQUENCE 274 AA; 30951 MW; D7C826BD3982592D CRC64;

Query Match 11.0%; Score 999; DB 10; Length 274;
Best Local Similarity 68.6%; Pred. No. 2.3e-59;
Matches 179; Conservative 38; Mismatches 42; Indels 2; Gaps 2;

Qy 1469 ESDIEDAYNEDDIPRIKINIEFGMTLREHME-RNWELOEGDMSKALVALHTTTISI 1527
Db 3 ECAIEDSFV-DOPEIPIKUNFEFTONLSYMOANNIEIEDADMSKALVAITPEVASI 61

Qy 1528 PTPKLNISRLRTEHQVYELPDSHRLDGMKREPDDPSVILLAIWTPGETANSQAQPEQ 1587
Db 62 PTPKLNVSRLRTEHQVYELPDSHPLLEGFNQREDDPCPYLLSIWTFGETAQSDAPKS 121

Qy 1588 KGGKASKMCFDETCSENSLRANSOTVRGTLIPORTAMRGSFPLNGTYFQVNELFA 1647
Db 122 VCSOENGELCASCTFCSCNSIREAQAKVRGTLIPORTAMRGSFPLNGTYFQVNEVFA 181

Qy 1648 DHESLKPIDVPRDWLWDLPRRTVYFGTSVTSIFRGLSTEQIOFCFWKGFVCGVGEQKT 1707
Db 182 DHDSRNPIDVPRSWIWLPRRTVYFGTSIPTIFKGLTTEIQHCFWRGFCVGRGFDRTS 241

Qy 1708 RAPRPLMARLHFPASKLNK 1728
Db 242 RAPRPLARLHFPASKITRNK 262

RESULT 7
O49498 PRELIMINARY; PRT; 917 AA.
ID O49498 AC
O49498; AC
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 106.3 KDA PROTEIN.
GN F28A23.180 OR AT4G34060.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RA Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,
RA Herzi A., Neumann S., Hohnsels J., Jesse T., Heijnen L., Vos P.,
RA Meves H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzi A.,
RA Neumann S., Meves H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021961; CAAL7566.1; -.
DR EMBL; AL161584; CAB80123.1; -.
DR InterPro; IPR003265; -.
DR Pfam; PF00730; HhH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
KW Hypothetical protein.
SQ SEQUENCE 917 AA; 106310 MW; 302B3770ACEIA454 CRC64;

Query Match 7.8%; Score 709; DB 10; Length 917;
Best Local Similarity 21.4%; Pred. No. 4.1e-39;
Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;

Qy 651 IAEIYRMNLYLGDKEREQONAMVLYKGALVYPSKRRKPRKVDIDDETRINWL 710
Db 388 IAKLIKDMGRLLKINKK-----VTTMIKADKKLV-----TAKVNLDPETIKEDWV 431

Qy 711 LMKGDEKEGDEEDKKKEKWEERVRFRGRADSFARIAMHLVQGDREPSKGSVDSV 770
Db 432 LM--VNDSPRSYDDKETEAKWKEREIFQTRIDLFINRMHRLQGNRKFKQKGSVDSV 489

Qy 771 IGVFLTQNVSHLSASFMSLAARPPKLSRSREDRNVRVSVVVEDPEGCILNLEIPSW 830
Db 490 VGVFLTQNTDTYLSNAPFSAVAKEP-----VDARGLSYVIEEPQ----- 530

Qy 831 QEKVQHPDSMEVSGVDSGSKQLRDCSNSGIERFNLEKSIQNLSEEVLSODSFDPAIF 890
Db 531 -----DAKSSSECI-----ILSDE----- 538

Qy 891 QSCGRVGSCKSDAEPTTRCETKTVSGTSQSVQTSQPNLSDELICLOGNERPHLYEGS 950
Db 539 ----- 543

Qy 951 GDVQKQETTNVAQKPDLEKTMNWKDSVCFQGPNDNTNQTTPSSSYEQCATRQPHVLDI 1010
Db 544 ----- 543

Qy 1011 EDFGNGQEGGLYSWMSISPRVDRVKNKNVPRFRFRQGGSVPREFTGQIIPSTPHELPGMG 1070
Db 544 ----- 543

Qy 1071 LSGSSSAVOEHODDTQHNOQDEMKNKASHLQKTFDLNLSSECLTRQSTKONTGCLP 1130
Db 544 ---SISKVEDHEN-----TAKRKNEXTGII- 565

Qy 1131 RDTAEDVVDPLSNNSLQNLIVESNSSNKEBTAVEYKETNATILREMKGTLDCKKPTS 1190
Db 566 ----EDEIVD----- 571

Qy 1191 QWDSLRKQVEGNEGRQERNKNMDSIDYEAIRRASISEISEAIKERMNMNLAIRIKDFL 1250
Db 572 -WNILRR-MYTKEG--SRPEMHMDSVNMDSVRLSQNVLETITIKRGQFRILS----- 620

Qy 1251 ERIVKHGGIDLEWLRSPDPKADKYLISIRGLGLKSVCEVRLLTFLHNLAFVDPNVGRI 1310
Db 621 ERIL----- 632

Qy 1311 AVRMGWVPLQPLPSLQHLLELFPVLESIOKFLWPLCKLDQRTLYELHQLITFGKVF 1370
Db 633 AVRLGLVLEPLPVGQMH-----QLFEYELHYQMITFGKVF 669
```



```
Db 1500 SOTESKPNLSLLPPKQYL-GLKLTSSNGLKLPKLPNRRNONQAKVEKET- 1554
QY 1175 LREMGKTLADGKKPTSQWDSLRKDVEGNGROE- ---RNKNMDSIDYEAIRRASISETS 1230
Db 1555 -----KEYKAKEPKKELGFKSDFKIEGASFKGINLSSGKPF- 1591
QY 1231 EAIKRGNNMLAVRIKDFLEIRIVXDHGIDLEWLRSPDKA-KDYLL-----SI 1280
Db 1592 EYPSGKSFDGMTKQVYKAEENKPKFSEFNLNLNNKTPLSNAPKPIALSNDGKKPSL 1651
QY 1281 RGLGLKSVCEVRLTLHLNLAFFVDNMGRIAVRMGVLPLOLPESLOLHLELYPVLES 1340
Db 1652 AGLNISKSSLASVSINALDKSGD-DVPKLAUSIDKLRKEPPPP-----PTPPAV 1700
QY 1341 QKFLWPLRC- ---KLDORTLYELHYLIITGKVFCTKSRPNCNAC-PMRGECRHFAAYAS 1396
Db 1701 TK- ---PALATEKEIRESSIADV-TPLAMITKINSAK-KPDFSALKPNLGNKQSSSSNEQ 1755
QY 1397 ARALAPAPERSLTSATDPVPPESSPPVPAIMIELPLPLEKSLAG- ---APSRENC 1450
Db 1756 KPLNPNFSALKLPKOKSEETKPNLSSLLPPKQTLGLKLPKLTSSNGLKLPKLPNQNQ 1815
QY 1451 EPIIEE- ---PASPGQECTEITESDIEDAYN-----EDPD----- 1482
Db 1816 NOEVKESKVKAKEPKKELGFKSDFKIEGASFKGINLSSGKFEVPSGKSFDDMIKOKA 1875
QY 1483 --EIPITKILNIQFGMTLREHMERNEMLQ-----EGDMSKALVALHPTTTSIPTP 1530
Db 1876 KTEQPKTEBNEPKFNFNLNLNNKTPLSNAPKPIALSLSNDSNKGQVOCALITIRPIP- 1933
QY 1531 KUKNISRLRTEHQVVELPDSHLLDGMKREDDPSYLLAIWTPGETANSQAQPEKQCG 1590
Db 1934 -LSKGDKPRADQSLKLGKS---LSNPTONKPD-----LSSLKPPKOKSEE 1974
QY 1591 GRAS 1594
Db 1975 NKPS 1978
```

```
RESULT 9
ID 076891 PRELIMINARY; PRT: 5327 AA.
AC 076891;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE EG:49E4.1 PROTEIN.
GN FUTSCH OR EG:49E4.1 OR CG3064.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RA Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.;
RT "Sequencing the distal x chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031128; CAA20006.1; -.
DR FlyBase; FBgn0025392; futsch.
SQ SEQUENCE 5327 AA; 575942 MW; FEFEE23A118FF38A CRC64;
```

```
Query Match 3.1%; Score 278; DB 5; Length 5327;
Best Local Similarity 19.7%; Pred. No. 4.8e-09;
Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;
QY 2 QSIMOSSAVNATEAQNDGSRQDVLFDLNTPOQKSKRRKKMPKVVVVEG-----K 55
```

```
Db 1986 ESIKD-----EAESKESRRE-----SVAESPLPSKEASRPASVAESIKDEAK 2031
QY 56 PKRKRKPAELPKVVVEGPKRKP-----RKAATQEKVKSKETSAKKNNLKESATKK 108
Db 2032 SKESRRESVAESKPLPSKEASRPASVAESIKDEAKESRRESAESAESPLPSKEASR 2091
QY 109 PANVG-----DMNKSPEVTLKSCRKALNFDLENPGDAR---QGDSE---SEIVQNSSGANSF 160
Db 2092 PASVAESVDEADKKEESRRE-----SMAESGKAQSIKGDQSPKLEVSRPESVAESV 2144
QY 161 SEIRDAIGTNGSFLDSVSQIDKTNCLGAMNOPLEVSMGNQDPKLTGAKLARDQDPLL 220
Db 2145 KD--DPVKESPSRRESVAGSVTADSDQSPLESKASRPESVSVDSVKDEAKESR 2202
QY 221 TRNQCOQFVATQNTQFPMENQOQWLMQKNOLIGFPNQOQPRMTIRNOQPCLAMNOOP 280
Db 2203 ESKTESVIPKAKDKSPKEVLQ-----PVSMT-----TIREDA-----DQP 2240
QY 281 MYLIGTPRALVSGNQ-----LGGPGNKRPIFLNHOTCLPAGNQLYSGPTDMHQLV 333
Db 2241 M-----KPSAERRESRESIAESIKASSPRDEKSPLASKEASRPGSVAESIKYDLDPQII 2294
QY 334 MSTGGQOH-----GLLKN-----QPGSLIRGOQPCVPLID---QOPATPK-----GF 374
Db 2295 KDKSTEHSRRESLEDKSAVTSEKSVSRPLSVASDHEAAVAIEDDAKSISPKDKSRPGF 2354
QY 375 THLNQVATSMSPGLRPHRSQSQVPTTYLHVESVR-----ILNGTTGTQORSRAPVD 428
Db 2355 -----VAETVSP-----IEATMEFSKIEVVEKSSLSLLOGSGGKLQTDSSPV-- 2400
QY 429 SLOODTHOQNYLILSHEISNGNGCKKALPONSLSPTPIMAKLEEAARGSKRQYHRAMQTE 488
Db 2401 ---DVAEGD---FSAVA-----SVSTVPTLTTPAELA-----QIGAATV 2436
QY 489 KHDNLAAQIAQS-QDVERHNSSTCVVEYLDAAKTKIKVVOENLHGMPEVTEIEDD-- 545
Db 2437 SSPLDEALRTPSAPEHISRADSPAECASEIASQDKSPQVLKES---SRPAWAEKDDAA 2494
QY 546 ---PTDGARKGKNTASISKAGKGN---SSPVKKT-----AEKEKIVPKTPARKGRA 592
Db 2495 QLKSSVEDLRSVPASTEISRPASAGETASSPIEEAPKDFAEFEQAEKAVLPLTELKGNL 2554
QY 593 GRKKSVPVPAHASEIQLWQPTPKPLSRKPKGKRKSIQDSGKARGSGELLCOOSTA 652
Db 2555 PTLSSPVDVAHAS-VQPAELSKVDIEKTASPIDEASPIDEASPIDEASPIDEASPIDEAS 2613
QY 653 EIIYRMQNLVLDGKEREQEQNAWL-----YKGDGALVPVYESKKRKPVDIIDEIT 705
Db 2614 ESV-----EKSADSRPPSVVESTKADSTKGDISPSP-ESVLEGPDDVVEKSKES 2663
QY 706 RIWNLIMGKDEKEGEDEKDKKKEKWEERVRFRGRADSFITARMHLVQGRFRSPWKS 765
Db 2664 RPPSV---SASITGDSKDVSR-----PAS 2685
QY 766 VYDSVIGVFLTONVSDHLSSAFMSLAARPPKLSRREDERNVRVSVVDEPEGCILNLN 825
Db 2686 VYESV-----KDEH-----DKAESRRESIAKVESVI---DEAG----- 2715
QY 826 EIPSMQEKVQHPSDMEVSGVDSGSKLEOLRDCSNSGIEREFNLEKSIQNL-----EVL 879
Db 2716 -----KSDSKSSQDSQKDEKSTLASKAESRRESVSVSSKDDAKESRPSVI 2764
QY 880 SS-----QDSFDPALFQSCGRVGC-----SCSKSDAEFPPTTRCTKTSGTS 922
Db 2765 ASGEVPVRESKPLDSKDTSRPGSVVESVTADEKSEQSRRESVAESYKADTKK-DGKS 2823
QY 923 QSVQTCSPNLSDEICLOGNE-----RPHLYEGSDVQKQETTNVAOKKPDLEKTMNKDSV 978
Db 2824 Q--EASRPSSVDELLKDDDEKQESRRQSIITGSHKAMSTMGDESPMDKADKESRPSV 2881
QY 979 CFGQPNDTNWTOTSSSVEQCATRQPHVLD-IEDFGMOGEGGLGYSWMSIS-PR--VDRV 1034
```

Db 2882 AESIKHENTKDEESPLGS-----RRDSVAESIKSDITKGEKSLPVSKEVSRPESVVGSI 2935
Qy 1035 KKNVPRFRFGGSGVPRFTGQII---PSTPHLPGLMGLSGSSAVQEHQDDTHQNOQ 1091
Db 2936 KDEAESRRRESVAESPESKDATSAPPSKEHSRP-----ESVLGSLKD 2980
Qy 1092 EMKASHLOKTFLLDNLSSBECTLRQSTK-----QNTDGLPRDRTA-----EDVVD 1140
Db 2981 EGDKTSRRVSADSIKDEKSLLSVQESAPSESAESLKDAAAPSQTSRPESVTSVKD 3040
Qy 1141 PLSNSSLQNLIVESNSNKEQTAVYKFNATILREMK-GTLADGKKPTSQWDSLRKDV 1199
Db 3041 GKSPVASKASRPASVAENAKDSADESQRPESLPQSKAGSIKDEKSLASKDAEAKSK 3100
Qy 1200 EGNREGROERNKNNMDSIDYAIRASTSEI--SEAIRKGMNNML-----AVRIK 1247
Db 3101 E--ESRRESVAEPPLVSKEVSRPASVAESVKDEAKSKEESPLMSKEASRPASVAGSVK 3158
Qy 1248 DFLERIYKDHGGIDLEWLRESPPDKA 1273
Db 3159 DEAEKSKE-----ESRRESVAEKS 3177
RESULT 10
Q9HCY0 PRELIMINARY; PRT; 2254 AA.
AC Q9HCY0
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR-LIKE NUCLEAR REGULATOR.
GN TFNR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kelter A.R., Herchenbach J., Wirth B.;
RT "The transcription factor like nuclear regulator (TFNR) contains a
RT novel 55-amino acid motif repeated 9 times and maps closely to SMNL."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ238520; CAC04245.1; -
SQ SEQUENCE 2254 AA; 252844 MW; F350E96F53F04CFE CRC64;

Query Match 2.7%; Score 245; DB 4; Length 2254;
Best Local Similarity 18.7%; Pred. No. 2.3e-07;
Matches 358; Conservative 250; Mismatches 722; Indels 584; Gaps 85;

Qy 11 NATEATEONDGSRD-----VLEFDLNTPOOK---PSKRRKFKMPKVYVEGKPKRK 59
Db 339 NKFREKTNWRIDKAFQEKRPFDFFFAHLQKVLAEBEKKQKSVKNHLSKEKSKTK 398
Qy 60 KRPAELPKVYVEG-----KPKRKPKAAT----- 84
Db 399 PRKNVKKVYACGVNNDPDESMSRISDTSQKQATVEEESLTLSDREDAEQVALEVD 458
Qy 85 --QEKVSKETGSAK---KNLKESATKK--PANVGDMNK-----SPEVLTKSCRK---- 129
Db 459 LNKRRRRKQDGNELGVNLLNENATVQAGPSGKHKKNQCAIRPELKEGEGCSKQML 518
Qy 130 --ALNFD-----LENPGDARQDSESEIVQNSS-----GANSFS 161
Db 519 SCTQNDIGVGFASTEKEVKRTDPILSLNNQDATSVATESSESSTDLPSFEVGIRALC 578
Qy 162 EIRDAIGG----TNGSFLDSVQIDKTNGL-----GAMNOPLEVSMGNQDKLSTGAKLA 212
Db 579 EVNNAEGSCIEERNVDLKNLSLEIDQENVKPMLRGFRP-----KPNLSRAGKSV 631
Qy 213 RDQ--QPDLLFRNOCQFPVATQNTQTFPMENQQAQLQMKNLQIFGPFQGNQPRWTIRNQ 271
Db 632 LSOGKTESKNSKSHKTSVBKNHVEKDKMNTLDILRMETTERENP-----EAEVTSVLGEKN 688

Qy 272 CLAMGNOOPMYLIGTPRALVSGNQOLGGPOGNK-----RPIFLN-----HOTC 315
Db 689 CLOEGSQ-----LKALRPVQVRGLQPKPNAGKAERKEILISQEEIGANVEKNENESC 743
Qy 316 LPAGNQLYSPTDMHQLVMSTGGQOGLLIKNOQPGSLIRGQBP-----CVPLIDQQ 367
Db 744 A-----DRDTQPMHEDQSRKDFEEDVILQPEKNDQFQNVQDEPKVLNELSVQENN 796
Qy 368 PATPKGFTHLNOMVATSMSSPGLRPHSQSOVPTYLHVESVRLNLTGTCQSRAPAY 427
Db 797 KA-----NKLNV-----PILTRFRQPKP-----NIGRTGRREISKE 831
Qy 428 DLSQODTHQGNKYILSHEISNGNGCKKALPQNSLPTPIMAKLEEARSKRYRRAQGT 487
Db 832 EVLEKILVSGEMAAALRETVLDTSPKEM-----VPAINTK-EMQSDLKETGTRRAISPR 885
Qy 488 EKHDNLAAQIAQSQDVERHNSSTCV-----EYLDAAKKTKIQKVQV-----NLHG 534
Db 886 EKILDVIDDTTEMETGLKAMGREICLREKTPREVDATE--EIDKLEEAGRREISPOKNG 943
Qy 535 MPPEVIEEDPTDGARKGKNTASISKGASGNSSPVKKTAEKCI--VPKTPAKGR- 591
Db 944 -PEEVKPLGEVETD-----LKATGNSSPREKTPGTDTEEDIDKNEETGRR 990
Qy 592 -----AGRKS-----VPPPAHAS-EIQLWQP-----TPPKTP 618
Db 991 KISPRENGPEVKPVDDEMETDLNATGREGSPREKTPREVDATEIDLEETEREVSPQENG 1050
Qy 619 LSRSPKPKGKRKSIQDSGK---ARGSGELLQCDQSIATIIYRMONLYLGDKERE--OEQN 673
Db 1051 LEEVKPLGEMETDLKATGRDFFRGKTPVI--DAIEI-----EIDLEETEREISPOEN 1103
Qy 674 AMVLYKDGAL---VPYESKRRPRKV-DIDDETTIWNLLMCKGDEKDEKDKKE 729
Db 1104 GLEEVKPLGEMQTLKATGREISPREKTPREVDATE-----EIDKLE 1146
Qy 730 KWEERRRFRGRADSFIAHMLVQGRDRFSPWKSGVVDVIGVFLTQNVSDHLSASFAM 789
Db 1147 E-----TGREISP-----EENGPEVKPVD 1169
Qy 790 SLAARFPKPLSSREDERNVSVVDEPEGILNLIPIPSQEKVQHPDSMEVSGVDSGS 849
Db 1170 ETDLKTTGREGSREKTR-----VIDAAEVETDLEE-----TEREISPOENGP 1214
Qy 850 KEQLRDCSNGIERFNFLEKSIONLEEVLSQDSFDPALFQSCRGVSCSCSKSDAEFP 909
Db 1215 EE-----VKPVGKMETDLKEIREEISQRE-----KVLAEFS 1245
Qy 910 TTR-----CET-----KTVSGTSQSV-----QTGSPNL-----SDEICLOGNER 943
Db 1246 AIREKEIDLKETGRDIPIMEKVSQKMAVVEEMADLKETKENFRERGSEECV----- 1300
Qy 944 PHLYEGSDYQKQFTTNVAQKPDLEKTMNWKSVCFQGPNDTNWOTTPSSSYEQCATR 1003
Db 1301 --TEEKVAELKQGTDISPRENELEETSTR-----QTOTHLMQSGNDSFAMPS- 1349
Qy 1004 QPHVLDIEDFGMOGEGLYGWSNISPRVDRVKNKNVPRP--FRQGG----SVPREFTGQ 1057
Db 1350 ----LDQNTSSE-----VLSMMHTPVVEERKNEKEVSSHFSHFKISSOTHESDKTEVOGI 1401
Qy 1058 IIPSTHELPGMGLSGSSSAVQEHQDDTHQNODEMKNASHLOKTFLLDNLSSBECTLRQ 1117
Db 1402 QSPDVPEQFSDINLSKSL-----POEQKPLEIKPAPFVRSRF----- 1438
Qy 1118 SSTQNITDGLPRDRTAED--VVDPLSNSSSLQNLIVESNSNKEQTAVYKFNATIL 1175
Db 1439 KRKPNLARAALKRETTSEKTYIEKSKTKMETIVQENNEQDTLPSQHDSEASLMS 1498
Qy 1176 REMKGTLDGCK-----PTSQW-----DSLKRDVEGNEGROERNKNNMDSI 1216
Db 1499 RE-KDTLIGHRNEEAVILPCTQTERNLSPNSCPKEESQAPVQKNDVSVVSGTNNNTF 1557

Qy	366	QOPATP----	KGFTHLNQMY----	ATMSSSEPL-----	RPHSQSQVPTTYLHVESVSRLN	413	
Db	775	QMEGSPNSSSFEHIA	SARSARDHAISLSE	PRMLWGSDPYPHAE	PQOATTPKATEPEDEV--	832	
Qy	414	GTWTCTORSFA-----	-----	-----	PAYDSLQODIHOGNKYILSHEI-----	446	
Db	833	-----	RSAAALDQOQITAA	YSVEHNOLEAHPKAD	FTRESSEAAQVKFLSRSEVEDVRPH	885	
Qy	447	-----	SNGCGCKALPONSS	LTPp-----	IMAKLEEARGSKRQYHRAMGOTEKHDLNLAQOIA	499	
Db	886	HTDANNQSACFEA-	PQKTLSPAQEERIS	AVESQPSRRKRSVSHGSHNTOKPDEORS	SPESA	944	
Qy	500	QSDQVHRHNSCTC	VEYLD-----	AAKTKIQQVVOENLHGMPE	VEIETHDDPTDQARGKN	555	
Db	945	GIPKV-----	TSRCIDSKEPIER	PEEKPKKEGFRS	SEGPKPEKVVKSKSETRWGPRPSSN	1000	
Qy	556	TASISKGASKGNS	SPYKK-----	-----	TAEK--EKCIVPKTPAKKGRAGR	594	
Db	1001	RREEVNDPRVRRSG	IKKPVLRDMKEER	QKEKEKAERTKVVV--	KPETEKTKDL	1058	
Qy	595	KKSVPPPAHAISE	IQLOWPMPKPTLS	RSKPKGK-----	GRKSIQDSQKARGPSGELLQODS	650	
Db	1059	PPPPPPQPPAPLO	POSVPPTQPEAK	FPSTETATLAQK	PSQTEKPLEPVSIVQVEPA	1118	
Qy	651	IAELIYPMQNL	YLGDKERQEQON	AMVLYKGDGALVY	ESKRRKPRP--KVYDIDTET---	705	
Db	1119	VKTV-----	NOOTMAAPVYK	BEKQPEKVISK-----	DLVIERPRPSRPAAVKKESTLPP	1167	
Qy	706	RTWNLMLMGKDE	KEGDEKDKKKEW	EEERRVFRGRAD-----	SFIARMLHVQ	754	
Db	1168	RTVW-----	-----	KEARDWFPDQYGR	GRGEYYSRGRSVYSGYGRGRGR	1212	
Qy	755	GDRRFSPW----	KGSYVDSVIGV	FLTQNVSDHLSSA	FMSLAARFPKPKLSSRSREDERNVR	810	
Db	1213	GHTRDYQYRDNK	PRAEHIPSGPL	QRESESTRSESS	DFEVV----PKRRRQRGSETD	1268	
Qy	811	SVVVE-----	DPEGCILNELI	PSWQEKVQHPSD	MEVSGVDSKQDLRDCNSGTERENFL	867	
Db	1269	SETHESASOKDS-	LSKGKLPKREER	PNKKPVK-----	PHSSFKPDNHV	1311	
Qy	868	EKSTQNLLEE	VLSSQDSFDPAI-----	-----	FQSCGR-VGSCSCSKSDAEFFPTR	912	
Db	1312	RIDNRLLERPY	VRDDBKAKPGL	PKEPTRRCRGDT	FRRGDPOGRPSRSTLRRPAYR	1371	
Qy	913	-CETKTVSGTSQ	SVQTSQPNLSDE-	ICLQGNRPHLY	EGSGDYQKQF-----	957	
Db	1372	DNQWNPQSEVP	KPEDGEPRRHEQ	FIADKRPKFERK	FDPAREPRRQRPTRPQR	1431	
Qy	958	-----	-----	TTNVAQ-----	KKPDLKTXMWNKD	976	
Db	1432	DKPPRFRRLRER	EAAKSNEVVAV	TGNTVNNVAQEP	YNTLGDISGNKTPDL-SNQNSSD	1490	
Qy	977	SVCFQOPRNDT	NNQOT--TPSSSV	EOCATOPHY----	LDIETDFCMOGEGLGYSWMSISPR--	1030	
Db	1491	QA-----	NEEWETAS	ESDNERERDEK	KNADLNAQTVVKGENVLPKREIARSF	1543	
Qy	1031	-----	VDRV-----	-----	KNKNVPRFFRQGSVPREFTGOIIP--	STPHELPGMWGL	1071
Db	1544	SSQRPVDRQNR	GNNGPPKSGRNF	SGPRNRRSG----	PPSKSGRGPFDQDQAGTTCVDL	1600	
Qy	1072	SGSSSAVQEHQ	DDTOHNOQDEN	KASHLOKTFDL	LLNSESCELTROSSTKQNTDGLCLR	1131	
Db	1601	INGSSA-----	HHQGVVNGTG--	OK-----	NSKDSGTKKREDPK-----	PG	1635
Qy	1132	DRTAEDVDP	PLSN-----	NSSLQILVIES	NSNKEQTAVYEKTNATILREMKGT	LADGKKP	1188
Db	1636	PKPKPEKVDAL	SQDLNNYASV	IID--DHEPVT	IEDPOSN--LNDQGFTEVVS	SKKO	1689
Qy	1189	TSOWDSLRK	QVNEGQR-----	-----	KNKNMDSIDYBAIRRAIS	EISELSEAIKERGNNM	1241
Db	1690	-----	QRLODEERRKKE	QEQVQWNNKANE-----	-----	KGROSOTSK	1724

RESULT 14

RESULTS	14
Q9PU36	PRELIMINARY; PRT; 5120 AA.
ID Q9PU36	
AC Q9PU36;	
DT 01-MAY-2000 (TrEMBLrel. 13, Created)	
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)	
DE ACZONIN (FRAGMENT).	
GN ACZ.	
OS Gallus gallus (Chicken).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC Gallus.	
OC NCBI_TaxID=9031;	
RN [1]	
RN SEQUENCE FROM N.A.	
RP TISSUE=BRAIN;	
RX MEDLINE=99439764; PubMed=10508862;	
RA Wang X., Kbschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,	
RA Kilianmann M.W.;	
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active	
RT zones, shares homology regions with rim and bassoon and binds	
RT profilin.";	
RL J. Cell Biol. 147:151-162(1999).	
DR EMBL; Y19187; CAB60725.1; -.	
DR HSP; P04410; IAZ5.	
DR InterPro: IPR000008; -.	
DR InterPro: IPR001478; -.	
DR InterPro: IPR001565; -.	
DR InterPro: IPR002965; -.	
DR Pfam; PF00168; C2; 2.	
DR Pfam; PF00595; PDZ; 1.	
DR PRINTS; PR00360; C2DOMAIN.	
DR PRINTS; PR01217; PRICHTEXTENS.	
DR PRINTS; PR00399; SYNAPTOTAGMN.	
DR PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_1.	
DR PROSITE; PSS0004; C2_DOMAIN_2; 2.	
DR SMART; SM00239; C2; 1.	
FT NON_TER 1	
SO SEQUENCE 5120 AA; 560760 MW; A658D9891B658412 CRC64.	

Query Match 2.6%; Score 238; DB 13; Length 5120;
Best Local Similarity 18.3%; Pred. No. 2.2e-06;
Matches 379; Conservative 247; Mismatches 689; Indels 760; Gaps 91;

QY 7 SSAVNATEQNDGSRDQVLEFDLNLKTPQOKPSKRKRKFMPKVVVEGKPKRKRK --- 62
DB 229 SQAPPTKPSLOQSGS-----VKQSQPAKQGGVPRPSAQAGPPKQOPGSEKPT 279
QY 63 -----PAELPKVVVEGK-----PKRKRKAATQ-----EKVKS 90
DB 280 AQOTGPAKOPPOPGKTPLOQTGVPKVPQAGPTKFSQTAGAKSILAQQGLTKPKG 339
QY 91 KETGSAKKNLKESATKP-----ANVGDM-----116
DB 340 QOPGPEKLOQQAOSTQVSTPKTKFCPLCTTTELLHTPEKANYNTCTOCHTVVCSL 399
QY 117 ---NKSPVT-----LKSC---RKALNFDLEN-----137
DB 400 CGFNPNHITKEKWLNCNQMORALGDLGSHGGPQGLPPPKQKTPPTASTAKPSQL 459
QY 138 -----PGDAQGDSESEIVQNS-----SGANSFSEIRDAIGGTNGSFELD 176
DB 460 OPGQKKDASPKPDPQQAQSKKPPVQKQPSMPGPPVKSQKTHAEPD-----TGQQID 514
QY 177 SV---SQDKTNGLAMQNP---LEVSMGNQDPKLTGAK--LARDQDQDLLTRNQOCQFP 229
DB 515 STPKSDOVKPTQAEKQKQPSIQKPTMDTPTTSAAPGVKQDLADPQSPS--TQOKVTDSP 572
QY 230 VATQNTQPPMENQ-----QAWLOMKNOILGFPFGNQPPRMTRNQOCPCLAMGN 277
DB 573 M-PETTKPPADTHPAGDKPDKPLQVSRQKSDPKLASOSGAKSADAKTKQKPEAPVXDD 631
QY 278 QOPMYLIGTPALVSGNQGLGPGCNRPPIPLNHTCLPAGNQLYGSPTDMHQLVMSTG 337
DB 632 PKLQTKPAPKPTKPAK---GPOAGTGPRTSAQA-POPOQKTPQOSRRSLNLG 687
QY 338 G-----QOGLLILNNQPGSLIIRQOQPCVPLIDQOP 368
DB 688 GITDAPKQPTTQETVTKGLFGFASIFSQASSLIISAGPGSQTSGPAPPATKQPP 747
QY 369 ATPKGFTHLQNVATSMSSPGLRPHSQOVPTTYLHVSVSRILNGTGTTCORSRAPAYD 428
DB 748 SOPPASQAPPEAAQAQPPPKAAPTKKTETP---LASEKLGPMASDSTLTTKGSDLEKKP 804
QY 429 SIQQDI-HQGNKYILSHSINGNCKKALPQNSLSPTPIMAKLEEARSGKQYHRAMGT 487
DB 805 SLAKDSKHQTAEKAPAELEQ---EKASQKVSCLC-----839
QY 488 EKHDNLQAQIAQSQDVERHNSSTCVELDAAKTKIKQVQENLHGM--PPEVIEIED- 544
DB 840 -KTGLNIGSK-----DPPNFNTC-----TECKKVVV-NLCGFNPMHIVEVQEW 881
QY 545 ---DPTDGARKG---NTASISGASKGNSPVKKTAKEKCIVPKTPAKKGRAGKK 596
DB 882 LCLNCQOTRAMSGQLDMGKVPLEK---LGSPQVSKP-----PATP-----QKQ 923
QY 597 SVPPPAHASEIQLMQTPPKTPSLRSKPK-----GKRKSI-----632
DB 924 PVPVSHSPQKSSSTPTTAAT-----KPEEPSVPEKVPKLOQGLKLETLSDAKTQQG 978
QY 633 -ODSKARGPSGELCQSDIAEIIYRMONLYLGDKEREQEQAAMVLYKGDGALVPYESSK 691
DB 979 KEDAKSKQGLFKFTPSADKIQVRSQKEDSRLOQTCLTKTPSSDKILHGVQKEDIKFQEA 1038
QY 692 RKPVRKVD-----IDDETRTNLWLMGDEKEGDEEDKKEKKEWEEERRVFRGRADSF 746
DB 1039 LAKTPSADKILHRQKQEDPKLQKMKMAKALSADKIQPEAQKEDVOLQVRLSKAVSADKI 1098
QY 747 IARMLVGDRDRFSPWKSVDVSVIGVFLTQNVSDHLSSSAFMSLAA-RFPKLSRSSRED 805
DB 1099 ---QHGIQKDLNLQHVKIEKTSVSEKIQEAQESKLOQDKLPKTLSEDKIPATVSS---1152

QY 806 ERNRSVVVEDEPGCILLNLNEIPSHQEKVQHPSDMEVSGVDSGSKEQLRDCNSGIERN 865
DB 1153 HKKLLSKSEDEKKPELL-----EKSTPHPKD-----KKEQI-----1183
QY 866 FLEKSIQNLLEEVLSSQSDFDPAIFQSGRGVSCSCSKSDAEFPTTRCETKTVSG--TSQ 923
DB 1184 -----TAETTHITQ 1194
QY 924 SVQTSPLNSDEICQGNRPHLYEGSDGVQKOE--TNNVAQKKPDLKTMNKDVSFCQ 982
DB 1195 KVEVEAP-----CDKLHEKK-----QEDVKKEDLTGTGIPQMVSKPEKAEKTPVPSR 1243
QY 983 -PRND-----TNWQITPSSSYEOCATRQPHVLDIEDFGHQGEGLGTSWMSISPR 1030
DB 1244 LPRSDEHVAEREKIEKEDKSDTSSQQO---KSPQ---GLSDTGYSSDGISSL-----1292
QY 1031 VDRVKNKVPFRFRQGGVPREFTGOIIPSTPHPELPGMLSGSSSAVQEHQDDQHNQO 1090
DB 1293 -----GEIPSH-----IPSEKDL-----REPSQK 1313
QY 1091 DEMNKAS-----HLQKTFDLNLSSEECITROSSTKQNTIDGCLPRDRTAEDVVDPL 1142
DB 1314 DTISQSPSPDLAKLESTVLSLEAQAASLTUDEKSVK-----KELYETY 1360
QY 1143 SNNSSLON-----ILVESNSNKEQTAVEYKETNATILREMKGTADGKKFTSQWDSL 1196
DB 1361 SEQTDQKHKTPLPVPESYSSEEDLEA-IOEGERTIAADSKG-----GASSQTDYKE 1413
QY 1197 KQVEGNE--GROERNKNMDSIDYE-----AIRASI-SEISEAIK-----ERGMNMLAV 1244
DB 1414 ED-GGNDTPARKQVDSVEDSSESENSPVPRKRKRASVSGSSSDEYKRRDSDSGSDE---1469
QY 1245 RIKDFLERIVKD-----HGGIDLEWL-----ESPDKAKDYLLSIRGL 1283
DB 1470 --EDFIRKOIEMSADEADAGSEDEDEFIRNOLKEISVTEOKKEVKSAKGTVGKHRM 1527
QY 1284 GLKSEVCEVRLTLHLNLPVDTNVGRIAVRMGW-----VPLQPLP-----ESL 1326
DB 1528 ARKS-----SAGYDEADAGR---RHSWHDDDDFTFDESPEPKYRETKSQDGEEL 1572
QY 1327 -----QLHLELYPVLESIQFLWPRCLCKLQORTLY-----ELHVQLITFGKVCT 1372
DB 1573 AISGGGLRRFTIENLSTITS-----KYSETPEQKGLIYFDEEPELEMESLT-----1621
QY 1373 KSRPNCNACPMRGECHIFASAYASARLALPAPEERSLTSATIPVPPSPPPVAIPMIELP 1432
DB 1622 -----DSPDRSRGEGSSSLHASSFTPGTSPTSVSS 1652
QY 1433 L-----PLEKSLASGAPSRENCE-----PIIEEPASQGECTEITESDIEDAYN 1478
DB 1653 LDESDSSPSHKLGGESKQOKARHSHGPLLTIED--SSEELREFEELKEQEKOR 1711
QY 1479 E-----DPDEI-----PIKLNIEQFGMTLREHMERNMELQEG 1511
DB 1712 ELEQOQRSSSKSKKDKDELRAQRRRERKPTPPSNLSDASPT--EELQAAEMEL- 1768
QY 1512 DMSKALVALHPHTTTSIPTPKLN-----IS---RLRTEHQVVELPDSHRL-----LDG 1556
DB 1769 -----LHRSSCSEYSPSIESDEGEFEISPEKIEIEQVKVYKLTAVSLYSPTEKLG 1820
QY 1557 MDKRPDDPSYLLAIWTPGETANSAPPEQKCGKASCKMCFDTECSECNLSREANSOT 1616
DB 1821 ALKEESGQ-----KTLKSAEEVYEMIKHTKSKSF-QIASEKDEVFE--RES 1865
QY 1617 VRGTLIIICPTAMGRSFFPLNGTYFQVNFELFADHESLAPIDVPRDWINDLPRRTVYFGTS 1676
DB 1866 LYGGMLI-----EDYIYESLIEDYINGTV 1889
QY 1677 VTSIF--RGLSTEQIQCFKFWKVCVGRFGEOKTRA 1709
DB 1890 DTNLMRODESENYIQ-----QKQKKEKIRA 1915

```

RESULT 15
Q9JLT1
ID Q9JLT1 PRELIMINARY; PRT; 4880 AA.
AC Q9JLT1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MULTIDOMAIN PRESYNAPTIC CYTOMATRIX PROTEIN PICCOLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a Presynaptic Zinc Finger Protein Structurally Related to
RT Bassoon.";
RL Neuron 25:203-214(2000).
DR EMBL; AF138789; AA07822.2; -
DR InterPro; IPR000008; -
DR InterPro; IPR001478; -
DR Pfam; PF00168; C2; 1.
DR PRINTS; PR01217; PRICEXTENS.
DR PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_1.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR SMART; SM00239; C2; 1.
KW Matrix protein.
SQ SEQUENCE 4880 AA; 530148 MW; 73951BE4ED83EA68 CRC64;

Query Match 2.68; Score 235.5; DB 11; Length 4880;
Best Local Similarity 19.38; Pred. No. 3e-06;
Matches 362; Conservative 239; Mismatches 666; Indels 609; Gaps 93;

QY 35 POQPSKRK---RKMPKVVEGKPKRKPAPLKPVEGVGPKPKPKKAATQEKVK 90
DB 612 PSQASPKKEPPSKDSEKALESKPP-EPKPPPE-PKKPPPK---KPPPLVKQPTLHG 666
QY 91 KETGSAKKNLKE-----SATKKP-----ANVGDMSNKSPVTLKSCRKALNFDLE 136
DB 667 PTPATAPQLPVAELPEAPAPPKPSGLPEQAKAPGVDEPKPKMT-----ETRADIQ 720
QY 137 NPGDARGDSESELVQNSSGANSFSETRDAIGTNGSFLDSVSOIDXTNGLGAMNQPLEV 196
DB 721 SSSSTKP-DILSSQVQSAQVKTASPLK-----TDSAKPSQSFPPTGKTTPLDS 769
QY 197 SMGNOPDKLSTGAKLARDQPDLLTRNOQCFPVATONTQPMENOQAWLQMKNLIGFP 256
DB 770 KAMPRP---ASDKIIISQPGSGESKDPKHIDPI--QKKDEPKAQ-----P 811
QY 257 FGNQOQPMTRNOQPCILAMNQOQPMYLLIGTPRA---LVSGNQOLGQPGQGNKRPFLN-- 311
DB 812 KGSKP-----ETKPVPKSGTPS---GT-RPTAGAAAPPQQPPKPPQSRRESNLIG 861
QY 312 -----HQTCPL---AGNOLY---SPTDMHQLVMSTGGQGHLLIKNQOQSLIRGQQ 358
DB 862 GITDAPKSPQTTTQETVTGTLFGFASIFSQANLISITAGQ-----GPH 906
QY 359 PCVLIDQOPATPKGFTHLNQMVATSSSGLRPHSQS---QVPTYLHVSVSRILGTT 416
DB 907 P-----QTGPAAP-----SKQAPTQSQAAGPAKSTQQLPPAPAKATAVKKEAKAA 955
QY 417 GTCQSR---APAYDSLQDIIHOGNKYILSHSNGCKKALPQNSSLPTPIAKLEEA 473
DB 956 AENLESPEQAAPTAKTEKD-----KKPPPAKVGVKPPP--SEPEKA 994
QY 474 RGSKRQYHRAMQTE-----KHDNLNLAQQAQSQDQDVERHNSSTCVYLDRAKTKIQ 525
DB 995 VPA-----HKPDKTKPKPACPLCRTELNLGSG-----EPPNFNTCTEC-----KNOVC 1038

```

```

QY 526 KVVQENLHGM--PPEVIEID-----DPTDGARKGK--NTASISKGASKGNSSPVKKTAE 576
DB 1039 -----NLCGFNPHTLHTEIOWELNCLNQTORAISSQGLDGMKMPAPSPGKASWPAPAE 1093
QY 577 KECIIVKTP-----AKGRAGRKKSVPPPAHASEIQLMQPTTPKPLSRKSPKGRK 630
DB 1094 PSS---QKTPTGTQVKGKKEAEKTEAEKPEKETAFTASIEKTPMVTVDQKLESESGKK 1150
QY 631 SIODSKARGPSGELLCCDSIAELIYRMQNLVGLDKREQEONAMVLYKGDGALVPYESK 690
DB 1151 SKVSA-----LPEKKPSSEKA-----ISADKK 1173
QY 691 KRKP-----RPKVIDDDETRIWNLLMGKDEKE----- 719
DB 1174 ERAPPAEKEPPEKPIPVDKLLPPEAKPLSSEGEKHEILKAHVQIPPEEPTGKVAAK 1233
QY 720 -GDEEK--DKKKE-----KMWEEERRVFRGRADSFIAHMLVQDGRFRSPWKGS 765
DB 1234 AGEEOQPDSPKPEALPGATPLTLPKAGEKERAVAQPAE---GSSKDGQGER--SKEK-- 1286
QY 766 VDSVIGVFLTONVSDHLSSSAFMSLAARPPKLSRSDREDERNVRVSVVEDPECCILN 825
DB 1287 -----TEKEEDKSDTSS--SQQPKSPQGLSDTYSS-----DGISGLG 1323
QY 826 EIPSWQEKVQHPSDMEVSGVDSGSKQOLRDCNSGIERFNFLEKS-----IQNLEEV 878
DB 1324 EIPSL-----IPSEKDLKK-----GLKDSFSQESSPSSPSDLAKLESTV 1364
QY 879 LSSQDSFDPAPFOSGVRGSCSCSKS-----DAEFTPTRCETKTVSGT----- 921
DB 1365 LSILEA-----QASTLVGEKAEKKTQPKISPEKPODQOQTATSETLDTITIEEIKE 1418
QY 922 SOSVQTGSPNLSDICLOG-----NERPHL-----YEGSGDVOKQETNVAOK 964
DB 1419 SQEKKV-SPKKSE---QGPFSRKEHKEKPELVDDLSPRRASVDSVDSSESNVPR 1474
QY 965 KPDLKTMNKKDSCVFCQGRNDTNWQTPSSSYEQCATRPHVLDIEDFGMQGEGLYSW 1024
DB 1475 K-----RRTSIGSSSDSEYKQ-----ED--SQSGSEEDF 1502
QY 1025 MS---ISPRVDRVKNKNVPRFFRQGGSVPREFTGQIIPSTPHELPMG-----LS 1072
DB 1503 IRKQIIEMSADEASGSEDEEFIR---SOLKEISGVGESQKREAEKKGKGVAGKHRRLT 1559
QY 1073 GSSSAVQEHODDTHQNOODEMNKASHLOKTFDLNLSSECLTFOSSTKON-----ITDG 1127
DB 1560 RKSTSFDDDDAGRRHSHWDEDD-----ETF-----DESPELKFKETKSQSEELVWAGG 1609
QY 1128 CLPRDRTAE---DVVDPLSNNSLQNLIVESNSNKBQTAVEYKETNATILREMGK--- 1180
DB 1610 GLRFRKTIENSTIADKYSSSQKTLIFYD---EEPELEMSLTDSPEDSRGSGSS 1665
QY 1181 -----TLADGKKPTSDSLRKDVGN-----EGROERNKNNN-----DSIDYE 1219
DB 1666 SLHASSFTPTGTSV--SLDESDSDSPSHKKGESKQKRAHRSHGHLPLTTIEDSSEE 1724
QY 1220 AIRRASISELSEALKERGMMNL-AVRKDFLRIYKDHGIDLEWLRESPPDKAKDYLL 1278
DB 1725 ELR-----EHEELKBOEKORELEOQOQSSKKSKKDKDELRAQRREPRKTPPS----- 1775
QY 1279 SIRGLGLKVECVRLTLTHLAFVDTNVGRIAVRMGVLPQLPESLQHLHLELYPVLE 1338
DB 1776 -----NLSPIEDAS-----PTELRQAAME----- 1796
QY 1339 SIOKFLWPRICKLDQRTLYELHYOLITFGKVFTKSRPNACPMRG-----ECRH 1389
DB 1797 -----ELH-----RSCSEYSPSIESDP-EGFEISPEKIIIEVQK 1829
QY 1390 FASAYASARLALPAPERSLTSATIPVPPSPFPVAIPMIELPLPLEKSLASCAPSNREN 1449
DB 1830 VYKLPAAVLSYPTDQSVNQKQEGVQKALSAEMEMQKPHKYK-----APPAANER 1884

```

QY 1450 CEPIIEEPASQOECTE-ITESDIEDAYNEDPDEIPTIKLNIEQGMGLREHMERNMEL 1508
Db 1885 DEVFEKEPLYGMLIEDYIYESLVEDTYNGSVGSLLT--ROEQNGFMQQRGQKQVL 1942
QY 1509 QE-----GDMSKALVALHPTTTSIPTKLNISRLRTEHQVYELPDShRLLD--- 1555
Db 1943 QEYIYDDPMQKISDLQKEFEYESLHSHVVPQEDIVSS-----YIIPESHEIVDLGS 1994
QY 1556 -----GMDKREPDDPSYLLAIWTPGETANSAOPPEQKGG--KASGKMCFD--ETCSEC 1606
Db 1995 MVMSTSEKKLLDADSAYEELMRQOVQVTDGSSPVQTTIGDDMAESTLDFDQVQDASLT 2054
QY 1607 NSLREANS---QTVRGTLIP--CRTAMRGSPPLNGTYF-----QVNELFADHES----- 1651
Db 2055 SSILSGASLTDSTSATLSIPDKITQOFSABEEDEYVTDYTREIQDIIA-HESLILTY 2113
QY 1652 -----SLKRPIDVP 1659
Db 2114 SEPSESATSVPPSDTP 2129

Search completed: July 5, 2001, 12:45:05
Job time: 167 sec

THIS PAGE BLANK (USPTO)